Expert Systems with Applications 38 (2011) 12249–12261

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Rule extraction based on granulation order in interval-valued fuzzy information system

Yi Cheng [a,b,*], Duoqian Miao [b]

[a] Department of Equipment and Engineering, Sichuan College of Architectural Technology, Deyang 618000, China
[b] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China

ARTICLE INFO

Keywords:
Granulation order
Interval-valued rough fuzzy sets
Interval-valued fuzzy information system
Rule extraction

ABSTRACT

Methods of fuzzy rule extraction based on rough set theory are rarely reported in incomplete interval-valued fuzzy information systems. This paper deals with such systems. Instead of obtaining rules by attribute reduction, which may have a negative effect on inducting good rules, the objective of this paper is to extract rules without computing attribute reducts. The data completeness of missing attribute values is first presented. Two different approximation methods are then defined. Two algorithms based on the two approximation methods, called MRBFA and MRBBA are proposed for rule extraction. The two algorithms are evaluated by a housing database from UCI. The experimental results show that MRBFA and MRBBA achieve better classification performances than the method based on attribute reduction.

1. Introduction

A basic issue in a rule-based system is extracting rules for classification or inference. Rules can be obtained from available data. Rough-set data analysis uses only internal knowledge, avoids external parameters, and does not rely on prior model assumptions such as probabilistic distribution in statistical methods, basic probability assignment in Dempster–Shafer theory. Its basic idea is to search for an optimal attribute set to generate rules through an objective knowledge induction process.

The classical rough set theory developed by Pawlak (1982) is used only to describe crisp sets. In order to describe a fuzzy concept in a crisp approximation space, Dubois and Prade (1987, 1990) extended the basic idea of rough sets and got a new model named rough fuzzy sets. This new model has been proven a promising tool for pattern recognition, data mining, and knowledge discovery (Asharafa & Murty, 2003, 2004; Bhatt & Gopal, 2005; Gong, Sun, & Chen, 2008; Greco, Inuiguchi, & Slowinski, 2006; Jiang, Wu, & Chen, 2005; Miao, Li, & Fan, 2005; Radzikowska & Kerre, 2002; Rajen & Gopal, 2005; Richard & Qiang, 2002; Sankar, 2004; Shen & Chouchoulas, 2002; Wang, 2003). There exist symbolic values, real values or interval values in a practical database (Richard & Qiang, 2002). For example, current, ID, temperature, time and voltage, such kinds of data are often described by interval values. However, the traditional rough fuzzy set theory cannot deal with these kinds of data effectively. Extending the rough fuzzy set theory of Dubois to a wider application is necessary. As a generalization of Zadeh's fuzzy set, the notion of interval-valued fuzzy sets was put forward for the first time by Gorzalczany (1988) and Turksen (1986). As to a fuzzy set, the interval-valued membership is easier to be determined than the single-valued one. Interval-valued fuzzy set theory has been applied to the fields of approximate inference, signal transmission, etc. Due to the complementarity between interval-valued fuzzy sets and rough sets, interval-valued rough fuzzy sets that combined interval-value fuzzy set with rough set was proposed. The definition of interval-valued rough fuzzy sets together with some important properties was put forward by Gong et al. (2008), a method of knowledge discovery was presented subsequently for interval-valued fuzzy information systems. The method classifies each object in a decision class according to its maximal membership represented by a fuzzy interval. However, suppose that the condition attribute set includes m attributes, then the antecedent of the rule must include m conditions, overfull conditions may reduce the classification accuracy and the applicability of the rules. Moreover, two memberships represented by fuzzy intervals are incomparable when one interval is nested in the other, then decision rules cannot be generated in this case. Aside from (Gong et al., 2008), few studies on fuzzy rule extraction are based on rough sets in interval-valued fuzzy information systems. It is necessary to establish a practical model for fuzzy rule extraction in interval-valued fuzzy information systems. The model should satisfy the following requirements: firstly the computational complexity of the model can be effectively reduced; secondly the applicability of extracted rules is preferable; thirdly rules can be generated when one interval is nested in the other.

* Corresponding author at: Department of Computer Science and Technology, Tongji University, Shanghai 201804, China.
E-mail address: chengyimail@sohu.com (Y. Cheng).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.04.003

The objective of attribute reduction is to reduce fuzzy attributes and learn fuzzy rules from fuzzy samples. Various attribute reduction methods have been proposed (Bhatt & Gopal, 2005; Deng, Chen, Xu, & Dai, 2007; Hu & Yu, 2004; Hu, Yu, & Xie, 2006, 2007; Jensen & Shen, 2004, 2007; Moshkov, Skowron, & Suraj, 2008; Shen & Jensen, 2004; Tsang, Chen, Tsang, Lee, & Yeung, 2005; Tsang, Chen, Yeung, Wang, & Lee, 2008; Wang, Ha, & Chen, 2005, 2007, 2008; Zhao, Tsang, Wang, Chen, & Yeung, 2006; Zhou & Wu, 2008) in rough fuzzy sets and fuzzy rough sets. In general, attribute reduction should be obtained before rule extraction (Pawlak, 1982). Since interval-valued rough fuzzy set theory is a generalization of traditional rough fuzzy set theory, it is natural to extract rules based on attribute reduction in such systems. However, this paper does not intend to obtain rules based on attribute reduction due to the following reasons. Usually, attribute reduction methods can be classified into three types: one based on the positive region (Bhatt & Gopal, 2005; Jensen & Shen, 2004, 2007; Shen & Jensen, 2004), one based on the discernibility matrix (Tsang et al., 2005, 2008; Wang et al., 2005; Zhang, Wu, Liang, & Li, 2001), and another based on entropy (Hu & Yu, 2004; Hu et al., 2006, Hu, Xie, & Yu, 2007). For example, Jensen and Shen (2007), Shen and Jensen (2004) conducted pioneering studies on attribute reduction based on a positive region and proposed an attribute reduction algorithm. An obvious limitation is that the algorithm may not be convergent on many real data sets or the selected attributes are unreliable. Moreover, the computational complexity of the algorithm often increases exponentially with the increase of input variable number and the data pattern size (Bhatt & Gopal, 2005). Bhatt and Gopal (2005) developed Shen's algorithm by improving the definition of the lower approximation operator on a compact computational domain. However, the dependency degree of a selected reduct may be larger than that of the entire attribute set due to the computing method of the positive region (Tsang et al., 2008). This is unreasonable according to the fact that more attributes will offer better approximations in a rough set framework (Tsang et al., 2008). Tsang et al. (2005, 2008) proposed an algorithm using a discernibility matrix to compute all attribute reducts. However, the computation complexity is NP-hard (Wang et al., 2005). Hu and Yu (2004, 2006) proposed an attribute reduction method based on information entropy. The attribute reduction concept is not constructed using existing fuzzy approximation operators (Yeung, Chen, Tsang, Lee, & Wang, 2005), and it is difficult to study the structure of attribute reduction (Zhao et al., 2006). Each attribute reduction method has its property. At the same time, each one has some flaws also. Therefore, rule extraction based on attribute reduction may be faulty sometimes. This paper intends to avoid the attribute reduction process and establish the structure of the approximation by introducing granulation order, then extract rules based on it.

From the viewpoint of granular computing,, a concept is depicted by both the upper and lower approximations under static granulation in the interval-valued rough fuzzy set theory defined by Gong et al. (2008). Provided the granulation is unchangeable, no matter whether the granulation is too fine or too coarse may be unacceptable. Excessively fine granulation may increase the time and cost, while an excessively coarse one may not satisfy requirements. We consider describing a concept under dynamic granulation. This means that a proper granulation family can be selected to describe a target concept according to the practical requirements. The notion of a granulation order was introduced in Liang, Qian, Chu, Li, and Wang (2005), Qian (2008). The objective of this paper is to find a novel method of rule extraction without computing attribute reduction in interval-valued fuzzy information systems. Based on the granulation order, two approximation methods in interval-valued rough fuzzy sets are proposed. A sequence of granulation spaces from coarse to fine can be obtained through adding one condition attribute at a time. Then the upper and lower approximations of forward approximation can be obtained under the given granulation order. Based on the forward approximation, a rule extraction algorithm named MRBFA is designed. Its characteristic is that the universe dwindles gradually and the approximation precision increases monotonously as the granulation order becomes longer. Thus the computational complexity of the algorithm can be effectively reduced. In the similar way, a sequence of granulation spaces from fine to coarse can be determined through deleting one condition attribute at a time. Then the upper and lower approximations of backward approximation are defined under the given granulation order. As an application of the backward approximation, an algorithm called MRBBA is proposed for decision-rule extraction from an interval-valued fuzzy decision table. The main characteristic of MRBBA is that much simpler rules can be extracted by keeping the approximation precision invariant.

The rest of this paper is organized as follows. Section 2 briefly introduces related notions of interval-valued fuzzy sets and interval-valued rough fuzzy sets. In Section 3, an algorithm is presented for data completeness in interval-valued fuzzy information systems. In Section 4, the forward approximation is proposed and important properties are obtained. A rule extraction algorithm called MRBFA based on the forward approximation is then designed, an example is illustrated. In Section 5, backward approximation is presented, and useful properties are deduced. For an interval-valued fuzzy set, its convergence degree is defined and proven to increase in a granulation order. A new rule extraction algorithm called MRBBA based on the backward approximation is proposed and illustrated. In Section 6, the performance of MRBFA and MRBBA are evaluated by a housing database from the UC Irvine Machine Learning Repository (UCI). Section 7 concludes the paper.

2. Preliminaries

In this section, for our further development, we briefly review the basic concepts of interval-valued fuzzy sets and interval-valued rough fuzzy sets.

2.1. Interval-valued fuzzy sets

Let I be a closed unit interval, i.e., $I = [0,1]$. Let $[I] = \{a = [a^-, a^+]: a^- \leqslant a^+,\ a^-,\ a^+ \in I\}$. For $\forall a \in I$, define $\bar{a} = [a,a]$, it is obvious that $a \in [I]$.

Definition 1. If $a_i \in [I]$, $i \in J$, $J = \{1,2,\ldots,m\}$, define

(1) $\bigvee_{i\in J} [a_i^-, a_i^+] = [\bigvee_{i\in J} a_i^-, \bigvee_{i\in J} a_i^+]$;
(2) $\bigwedge_{i\in J} [a_i^-, a_i^+] = [\bigwedge_{i\in J} a_i^-, \bigwedge_{i\in J} a_i^+]$;
(3) $[a_i^-, a_i^+]^c = [1 - a_i^+, 1 - a_i^-]$.
In particular, for $a_i \in [I]$, $i = 1, 2$, we define
(4) $a_1 = a_2 \iff a_1^- = a_2^-,\ a_1^+ = a_2^+$;
(5) $a_1 \leqslant a_2 \iff a_1^- \leqslant a_2^-,\ a_1^+ \leqslant a_2^+$;
(6) $a_1 < a_2 \Leftrightarrow a_1 \leqslant a_2,\ a_1 \neq a_2$;
(7) $a_1 \leqslant_w a_2 \iff a_1^- + a_1^+ \leqslant a_2^- + a_2^+,\ a_1^+ - a_1^- \leqslant a_2^+ - a_2^-$;
(8) $a_1 <_w a_2 \Leftrightarrow a_1 \leqslant_w a_2,\ a_1 \neq a_2$.

Definition 2. Let X be an ordinary non-empty set, the mapping A: $X \to [I]$ is called an interval-valued fuzzy set on X. The set of interval-valued fuzzy sets on X is denoted by $F^I(X)$.

Similar to fuzzy sets, the operators $\subseteq$, $\cap$, $\cup$, and the complement of interval-valued fuzzy sets are defined as follows. For $A, B \in F^I(X)$, $A \subseteq B$ means $A(x) \leqslant B(x)$ for $\forall x \in X$, $(A \cap B)(x) = \wedge\{A(x), B(x)\}$, $(A \cup B)(x) = \vee\{A(x), B(x)\}$, $(\sim A)(x) = 1 - A(x)$.

Definition 3. If $A \in F^I(X)$, let $A(x) = [A^-(x), A^+(x)]$, where $x \in X$, then two fuzzy sets $A^-: X \to I$, and $A^+: X \to I$ are called lower fuzzy set and upper fuzzy set about A, respectively.

2.2. Interval-valued rough fuzzy sets

Let U be a non-empty finite universe, R be an equivalence relation on$U$. $[x]_R$ is the equivalence class containing x.

Definition 4 Gong et al. (2008). For any interval-valued fuzzy set F, the lower and upper approximations of F about the approximation space (U,R) are defined as follows:

$$\mu_{\underline{apr}_R(F)}(x) = \inf_{y\in[x]_R} \mu_F(y) = [\mu_{\underline{apr}_R(F^-)}(x), \mu_{\underline{apr}_R(F^+)}(x)] = [\inf_{y\in[x]_R} \mu_{F^-}(y), \inf_{y\in[x]_R} \mu_{F^+}(y)],$$

$$\mu_{\overline{apr}_R(F)}(x) = \sup_{y\in[x]_R} \mu_F(y) = [\mu_{\overline{apr}_R(F^-)}(x), \mu_{\overline{apr}_R(F^+)}(x)] = [\sup_{y\in[x]_R} \mu_{F^-}(y), \sup_{y\in[x]_R} \mu_{F^+}(y)].$$

If for any $x \in U$, $\mu_{\underline{apr}_R(F)}(x) = \mu_{\overline{apr}_R(F)}(x)$, then the interval-valued fuzzy set F is definable about (U,R). Otherwise the interval-valued fuzzy set F is rough about (U,R), and F is called an interval-valued rough fuzzy set.

If F is an ordinary fuzzy set of universe U, then $F^- = F^+$. Therefore, the interval-valued rough fuzzy set degenerates into a classical rough fuzzy set.

3. Data completeness in interval-valued fuzzy information systems

Data completeness in interval-valued fuzzy information systems is the usual prerequisite for rule extraction. The process of converting an incomplete interval-valued fuzzy information system into a complete one, namely complementing the missing attribute values with specified values via some techniques, is called the completeness of incomplete interval-valued fuzzy information system. Multiple methods have been proposed for classical incomplete information systems (Grzymala-Busse & Hu, 2001; Grzymala-Busse & Grzymala-Busse, 2007; Kononenko, Bratko, & Roskar, 198; Qin, 2005; Wang, 2001; Zhu, Zhang, & Fu, 2004). A simple method is to either delete objects that miss their attribute values or replace their values with the most common one. The second method is based on probability statistics, e.g., Bayes method and multivariate linear regression analysis method. A Bayesian formula is used to determine the probability distribution of the missing value over the possible values (Kononenko et al., 198). This method either chooses the most likely value or divides the object into fractional objects, each with one possible value weighted according to the probabilities. In application, due to the vast state-space of data set, it's difficult to determine the probability distribution. Thus, the traditional statistical technique may not be the best choice. The third method is based on a classical rough sets theory, such as the rough set theory-based incomplete data analysis approach (ROUSTIDA) (Wang, 2001; Zhu et al., 2004). The basic principle is to make the missing attribute values of the objects are consistent to the ones of the other similar objects, i.e. to reduce the differences between attribute values as possible. For ROUSTIDA, the attribute differences between different objects are reflected by a discernibility matrix, and missing values are replaced with those of indiscernibly objects. We present the data completeness method based on ROUSTIDA in allusion to interval-valued fuzzy information system.

Let $S = (U, C \cup D)$ be an interval-valued fuzzy information system with decision attributes, we call S an interval-valued fuzzy decision table, where the condition attributes in C are crisp data, and the decision attributes in D are fuzzy interval numbers, U and C are denoted as: $U = \{x_1, x_2, \ldots, x_n\}$, $C = \{a_1, a_2, \ldots, a_m\}$ and $|U| = n$, $|C| = m$.

Definition 5. Let $S = (U, C \cup D)$ be an interval-valued fuzzy information system, its discernibility matrix is a $n \times n$ square matrix $M(C) = \{M(i,j)\}_{n\times n}$, $1 \leqslant i \leqslant n = |U|$. The matrix unit is defined as follows:

$$M(i,j) = \{k : (a_k(x_i) \neq a_k(x_j)) \text{ and } (a_k(x_i) \neq *) \text{ and } (a_k(x_j) \neq *)\},\quad k = 1,2,\ldots,m;\ i,j = 1,2,\ldots,n.$$

Definition 6. Let $S = (U, C \cup D)$ be an interval-valued fuzzy information system, missing attribute set MAS_i of object x_i, indiscernible object set NS_i for object x_i and the set of objects with missing attribute values MOS in interval-valued fuzzy information system S are defined as:

$$MAS_i = \{k : a_k(x_i) = *,\ k = 1,2,\ldots,m\},\quad NS_i = \{j : M(i,j) = \emptyset,\quad i \neq j,\ j = 1,2,\ldots,n\},$$

$$MOS = \{i : MAS_i \neq \emptyset,\quad i = 1,2,\ldots,n\}.$$

Because of multiple missing data and their different distributions, supplement of the missing data may not be achieved after a single computation of initial discernibility matrix. It may need repeatedly compute the discemibility matrix and analyze completeness. Many transient information systems are generated in the supplement process.

The initial interval-valued fuzzy information system specified as S^0, and the object set is $\{x_i^0\}$, the corresponding discernibility matrix is M^0, the missing attribute set of $x_i^0$is MAS_i^0, the indiscernible object set is NS_i^0. Let S^r be the information system after rth completeness analysis, and its object set is $\{x_i^r\}$, the corresponding discernibility matrix is M^r, the missing attribute set of $x_i^r$is MAS_i^r, its indiscernible object set is NS_i^r. A completeness algorithm for incomplete interval-valued fuzzy information system is given as follows:

Input: incomplete information system $S^0 = (U^0, C \cup D)$
Output: complete information system $S^r = (U^r, C \cup D)$
(1) Compute M^0, MAS_i^0 and MOS^0. Let $r = 0$;
(2) 2.1 For $\forall i \in MOS^r$, compute NS_i^r;
2.2 Generate S^{r+1}:
2.2.1 For $i \notin MOS^r$, let $a_k(x_i^{r+1}) = a_k(x_i^r)$, $k = 1,2,\ldots,m$;
2.2.2 For all $i \in MOS^r$, make loops for all $k \in MAS_i^r$:
2.2.2.1 If $|NS_i^r| = 1$, let $j \in NS_i^r$, if $a_k(x_j^r) = *$, then $a_k(x_i^{r+1}) = *$; otherwise $a_k(x_i^{r+1}) = a_k(x_j^r)$;
2.2.2.2 Otherwise
(i) If $\exists j_0 \in NS_i^r$ and $j_1 \in NS_i^r$, with the condition $(a_k(x_{j_0}^r) \neq *) \wedge (a_k(x_{j_1}^r) \neq *) \wedge (a_k(x_{j_1}^r) \neq a_k(x_{j_0}^r))$, then $a_k(x_i^{r+1}) = *$;
(ii) Otherwise, if $\exists j_0 \in NS_i^r$, with the condition $(a_k(x_{j_0}^r) \neq *)$, then $a_k(x_i^{r+1}) = a_k(x_{j_0}^r)$;
(iii) Otherwise, $a_k(x_i^{r+1}) = *$;
2.3 If $S^{r+1} = S^r$, go to (3); otherwise compute M^{r+1}, MAS_i^{r+1} and MOS^{r+1}; Let $r = r + 1$; go to (2);
(3) If there are still missing data in the information system, combination completeness approach is adopted for further process;
(4) The end.

The computation complexity of Step 2.2 depends on the distribution and quantity of the missing data. Usually, the missing data comprise of only a small portion of total data, so the computation complexity is relative low. In Step 2.3, the corresponding discernibility matrix of information system M^{r+1} is needed, according to Definition 5, $mn(n-1)/2$ times of computation is required.

To illustrate the operation of the algorithm, an example is given here. The data set comes from Table 4.1 in Gong et al. (2008). In the data set, there are no missing data. The author generates randomly some missing values with certain ratio to get an incomplete information system denoted as Table 1. Table 1 is an incomplete interval-valued fuzzy information system with 5 missing data, where $U=\{x_1,x_2,\ldots,x_{10}\}$ is a set of objects, $C=\{a_1,a_2,a_3\}$ is a condition attribute set, the decision attribute d is fuzzy, separated into three linguistic terms F_1, F_2, F_3 and F_1, F_2, F_3 are interval-valued fuzzy sets. Missing data are represented by *.

According to the algorithm, the completed interval-valued fuzzy information system S^1 is shown in Table 2. It can be seen from Table 2, all the missing data are completed after one time of computation.

By comparing Table 2 with Table 4.1 in Gong et al. (2008), one can easy see that Table 2 is the same as Table 4.1 in Gong et al. (2008). The example shows that the algorithm fully utilize the laws suggested by the data in information systems, and it can easily process completeness analysis of an incomplete information system, thus it can be adopted as a pretreatment method in data mining.

Table 1
An incomplete interval-valued fuzzy information system S^0.

U	a_1	a_2	a_3	F_1	F_2	F_3
x_1	2	1	3	[0.7, 0.9]	[0.15, 0.2]	[0.4, 0.5]
x_2	3	2	1	[0.3, 0.5]	[0.5, 0.7]	[0.35, 0.4]
x_3	*	1	3	[0.7, 0.8]	[0.3, 0.4]	[0.1, 0.2]
x_4	2	2	3	[0.15, 0.2]	[0.5, 0.8]	[0.2, 0.3]
x_5	1	*	4	[0.05, 0.1]	[0.2, 0.3]	[0.65, 0.9]
x_6	1	1	2	[0.1, 0.2]	[0.35, 0.5]	[1.0, 1.0]
x_7	*	2	1	[0.25, 0.4]	[1.0, 1.0]	[0.3, 0.4]
x_8	1	1	4	[0.1, 0.2]	[0.25, 0.4]	[0.5, 0.6]
x_9	2	1	*	[0.45, 0.7]	[0.25, 0.3]	[0.2, 0.3]
x_{10}	3	*	1	[0.05, 0.1]	[0.8, 0.9]	[0.05, 0.2]

Table 2
A completed interval-valued fuzzy information system S^1.

U	a_1	a_2	a_3	F_1	F_2	F_3
x_1	2	1	3	[0.7, 0.9]	[0.15, 0.2]	[0.4, 0.5]
x_2	3	2	1	[0.3, 0.5]	[0.5, 0.7]	[0.35, 0.4]
x_3	2	1	3	[0.7, 0.8]	[0.3, 0.4]	[0.1, 0.2]
x_4	2	2	3	[0.15, 0.2]	[0.5, 0.8]	[0.2, 0.3]
x_5	1	1	4	[0.05, 0.1]	[0.2, 0.3]	[0.65, 0.9]
x_6	1	1	2	[0.1, 0.2]	[0.35, 0.5]	[1.0, 1.0]
x_7	3	2	1	[0.25, 0.4]	[1.0, 1.0]	[0.3, 0.4]
x_8	1	1	4	[0.1, 0.2]	[0.25, 0.4]	[0.5, 0.6]
x_9	2	1	3	[0.45, 0.7]	[0.25, 0.3]	[0.2, 0.3]
x_{10}	3	2	1	[0.05, 0.1]	[0.8, 0.9]	[0.05, 0.2]

4. Forward approximation in interval-valued rough fuzzy sets

4.1. The concept of forward approximation

In interval-valued rough fuzzy sets (IVRF) theory defined by Gong et al. (2008), the concept is described by the upper and lower approximations under static granulation. However, the fixed granulation may limit the application of IVRF. A sequence of granulation spaces from coarse to fine can be obtained through adding one condition attribute at a time. Then the upper and lower approximations of a target concept under the given granulation order can be obtained.

The notion of granulation order was introduced by Liang et al. (2005), Qian (2008): let $S=(U,A)$ be an information system, where U is a non-empty set of finite objects (the universe), A is a non-empty finite set of attributes and $P,Q\in 2^A$. Define a partial relation $\prec$ as follows: $P\prec Q$ ($Q\succ P$) if and only if, for every $P_i\in U/P$, there exists $Q_j\in U/Q$ such that $P_i\subseteq Q_j$, where $U/P=\{P_1,P_2,\ldots,P_m\}$ and $U/Q=\{Q_1,Q_2,\ldots,Q_n\}$ are equivalence classes induced by P and Q.

A partition induced by an equivalence relation provides a granulation space for describing a target concept. Let $R_i\in 2^A$ $(i=1,2,\ldots,n)$ be a family of equivalence relations with $R_1\succ R_2\succ\cdots\succ R_n(R_1\prec R_2\prec\cdots\prec R_n)$, the sequence of granulation spaces from coarse to fine (from fine to coarse) determined by $R_i\in 2^A$ $(i=1,2,\ldots,n)$ is named a granulation order. The upper and lower approximations of forward approximation under a granulation order are defined as follows.

Definition 7. Let $S=(U,A)$ be an information system, F be an interval-valued fuzzy set of U and $P=\{R_1,R_2,\ldots,R_n\}$ be a family of attribute sets with $R_1\succ R_2\succ\cdots\succ R_n(R_i\in 2^A, i=1,2\ldots,n)$, P-upper approximation $\overline{apr}_P(F)$ and P-lower approximation $\underline{apr}_P(F)$ of forward approximation of F are defined as:

$$\mu_{\overline{apr}_P(F)}(x)=[\mu_{\overline{apr}_P(F^-)}(x),\mu_{\overline{apr}_P(F^+)}(x)]=[\sup_{x\in[x]_{R_n}}\mu_{F^-}(x),\sup_{x\in[x]_{R_n}}\mu_{F^+}(x)],$$

$$\mu_{\underline{apr}_P(F)}(x)=[\mu_{\underline{apr}_P(F^-)}(x),\mu_{\underline{apr}_P(F^+)}(x)] =\begin{cases}[\inf_{x\in U_1/R_1}\mu_{F^-}(x),\inf_{x\in U_1/R_1}\mu_{F^+}(x)], & x\in W_1\\ \cdots & \\ [\inf_{x\in U_n/R_n}\mu_{F^-}(x),\inf_{x\in U_n/R_n}\mu_{F^+}(x)], & x\in W_n\\ [\inf_{x\in U_{n+1}/R_n}\mu_{F^-}(x),\inf_{x\in U_{n+1}/R_n}\mu_{F^+}(x)], & x\in U_{n+1}\end{cases},$$

where $U_1=U, U_i=U_{i-1}-W_{i-1}$ $(i=2,3,\ldots,n+1)$, $W_{i-1}=\{x|\mu_{\underline{apr}_{R_{i-1}}(F)}(x)=[\inf_{x\in U_{i-1}/R_{i-1}}\mu_{F^-}(x),\inf_{x\in U_{i-1}/R_{i-1}}\mu_{F^+}(x)]\geqslant[\eta^-,\eta^+]\}=\{x|\inf_{x\in U_{i-1}/R_{i-1}}\mu_{F^-}(x)\geqslant\eta^-,\inf_{x\in U_{i-1}/R_{i-1}}\mu_{F^+}(x)\geqslant\eta^+\}$, where η^-, $\eta^+\in[0.5,1]$ and $\eta=[\eta^-,\eta^+]\in[I]$ is a suitable threshold.

The boundary $BN_P(F)$ of F is defined as:

$$\mu_{BN_P(F)}(x)=[\sup_{y\in[x]_{R_n}}\mu_{F^-}(y)\wedge(1-\inf_{y\in[x]_{R_n}}\mu_{F^+}(y)),\sup_{y\in[x]_{R_n}}\mu_{F^+}(y)\wedge(1-\inf_{y\in[x]_{R_n}}\mu_{F^-}(y))].$$

Remark 1. The main idea of Definition 7 is that in the coarsest granulation space decided by R_1, delete those objects whose decision class can be determined and obtain updated universe U_1-W_1. In the coarser granulation space determined by adding a condition attribute, for the updated universe U_1-W_1, delete the objects whose decision class can be determined and update the universe again. This process is repeated until the updated universe becomes an empty set or there is no condition attribute can be added. Although approximation operators are equivalent to the ones in Definition 4, the structure of approximation operators in Definition 7 reflects the granulation spaces changing from coarse to fine. Definition 7 shows that the universe dwindles as the granulation space becomes fine. This helps reduce computational complexity.

Remark 2. Upper and lower approximations are not symmetrical. In many applications, computing the upper approximation is not always necessary. For simplicity, the upper approximation operator is not represented in a structural form.

Definition 7 shows that a target fuzzy concept is approached by the upper approximation $\overline{apr}_P(F)$ and variable lower approximation $\underline{apr}_P(F)$. In the process of approximate classification, there usually exists incompatibility between the approximate classification result and the decision class due to the unavoidable roughness of problem description. The closer the lower approximation $\underline{apr}_P(F)$ is to F, the higher the compatibility between the approximate classification result and the decision class. The result of the forward approximation is that the universe is decomposed into a union of several subsets, i.e., $U = W_1 \cup W_2 \cup \cdots W_n \cup U_{n+1}$. Each subset is located in different granulation levels, and it is the maximal subset satisfies the given threshold in the corresponding granulation. The number of "∪" quantitatively reflects the compatible extent between the approximate classification result and decision class.

Theorem 1. *Let $S = (U,A)$ be an information system, F be an interval-valued fuzzy set of U and $P = \{R_1,R_2,\ldots,R_n\}$ be a family of attribute sets with $R_1 \succ R_2 \succ \cdots \succ R_n (R_i \in 2^A, i = 1,2\ldots,n)$. Let $P_i = \{R_1,R_2,\ldots,R_i\}$, then for $\forall P_i$, $(i = 1,2,\ldots,n)$, the following properties hold:*

$$\underline{apr}_P(F) \subseteq F \subseteq \overline{apr}_P(F), \tag{1}$$

$$\underline{apr}_{P_1}(F) \subseteq \underline{apr}_{P_2}(F) \subseteq \cdots \subseteq \underline{apr}_{P_n}(F), \tag{2}$$

$$BN_{P_1}(F) \supseteq BN_{P_2}(F) \supseteq \cdots \supseteq BN_{P_n}(F). \tag{3}$$

Proof. It should be noted that $\mu_{\underline{apr}_P(F)}(x) = \mu_{\underline{apr}_{R_n}(F)}(x)$ in essence. Then $\mu_{\underline{apr}_P(F)}(x) = \mu_{\underline{apr}_{R_n}(F)}(x) \leqslant \mu_F(x) \leqslant \mu_{\overline{apr}_{R_n}(F)}(x) = \mu_{\overline{apr}_P(F)}(x)$, for $\forall x \in U$. That is $\underline{apr}_P(F) \subseteq F \subseteq \overline{apr}_P(F)$.

To prove (2), we prove $\underline{apr}_{P_1}(F) \subseteq \underline{apr}_{P_2}(F)$ firstly.

$$\mu_{\underline{apr}_{P_1}(F)}(x) = \begin{cases} [\inf_{x\in U_1/R_1}\mu_{F^-}(x), \inf_{x\in U_1/R_1}\mu_{F^+}(x)], & x \in W_1, \\ [\inf_{x\in U_2/R_1}\mu_{F^-}(x), \inf_{x\in U_2/R_1}\mu_{F^+}(x)], & x \in U_2 = U_1 - W_1, \end{cases}$$

$$\mu_{\underline{apr}_{P_2}(F)}(x) = \begin{cases} [\inf_{x\in U_1/R_1}\mu_{F^-}(x), \inf_{x\in U_1/R_1}\mu_{F^+}(x)], & x \in W_1, \\ [\inf_{x\in U_2/R_2}\mu_{F^-}(x), \inf_{x\in U_2/R_2}\mu_{F^+}(x)], & x \in W_2 \subseteq U_2 = U_1 - W_1, \\ [\inf_{x\in U_3/R_2}\mu_{F^-}(x), \inf_{x\in U_3/R_2}\mu_{F^+}(x)], & x \in U_3 = U_2 - W_2 = U_1 - W_1 - W_2. \end{cases}$$

Obviously, when $x \in W_1$, $\mu_{\underline{apr}_{P_1}(F)}(x) = \mu_{\underline{apr}_{P_2}(F)}(x)$; otherwise, $\mu_{\underline{apr}_{P_1}(F)}(x) \leqslant \mu_{\underline{apr}_{P_2}(F)}(x)$. That is, for $\forall x \in U$, $\mu_{\underline{apr}_{P_1}(F)}(x) \leqslant \mu_{\underline{apr}_{P_2}(F)}(x)$. Thus we can obtain that $\underline{apr}_{P_1}(F) \subseteq \underline{apr}_{P_2}(F)$. Similarly, we can show the other inequalities. Therefore, $\underline{apr}_{P_1}(F) \subseteq \underline{apr}_{P_2}(F) \subseteq \cdots \subseteq \underline{apr}_{P_n}(F)$.

To prove (3), we prove $BN_{P_1}(F) \supseteq BN_{P_2}(F)$ firstly.

$$\mu_{BN_{P_1}(F)}(x) = [\sup_{y\in[x]_{P_1}} \mu_{F^-}(y) \wedge (1 - \inf_{y\in[x]_{P_1}}\mu_{F^+}(y)), \sup_{y\in[x]_{P_1}} \mu_{F^+}(y) \wedge (1 - \inf_{y\in[x]_{P_1}}\mu_{F^-}(y))],$$

$$\mu_{BN_{P_2}(F)}(x) = [\sup_{y\in[x]_{P_2}} \mu_{F^-}(y) \wedge (1 - \inf_{y\in[x]_{P_2}}\mu_{F^+}(y)), \sup_{y\in[x]_{P_2}} \mu_{F^+}(y) \wedge (1 - \inf_{y\in[x]_{P_2}}\mu_{F^-}(y))].$$

It is clear that $[x]_{P_1} \supseteq [x]_{P_2}$, then $\sup_{y\in[x]_{P_1}}\mu_{F^-}(y) \geqslant \sup_{y\in[x]_{P_2}}\mu_{F^-}(y)$, $\inf_{y\in[x]_{P_1}}\mu_{F^+}(y) \leqslant \inf_{y\in[x]_{P_2}}\mu_{F^+}(y)$, therefore, $(\sup_{y\in[x]_{P_1}}\mu_{F^-}(y) \wedge (1 - \inf_{y\in[x]_{P_1}}\mu_{F^+}(y))) \geqslant (\sup_{y\in[x]_{P_2}}\mu_{F^-}(y) \wedge (1 - \inf_{y\in[x]_{P_2}}\mu_{F^+}(y))$; Similarly, we can show $(\sup_{y\in[x]_{P_1}}\mu_{F^+}(y) \wedge (1 - \inf_{y\in[x]_{P_1}}\mu_{F^-}(y))) \geqslant (\sup_{y\in[x]_{P_2}}\mu_{F^+}(y) \wedge (1 - \inf_{y\in[x]_{P_2}}\mu_{F^-}(y))$, Thus, $\mu_{BN_{P_1}(F)}(x) \geqslant \mu_{BN_{P_2}(F)}(x)$, that is $BN_{P_1}(F) \supseteq BN_{P_2}(F)$. Analogously, the other inequalities can be proven. □

Theorem 1 states that the lower approximation enlarges as the granulation order becomes longer by adding equivalence relations.

To describe the uncertainty of concepts under a granulation order, the approximation precision is defined as follows.

Definition 8. Let $S = (U,A)$ be an information system, F be an interval-valued fuzzy set of U and $P = \{R_1,R_2,\ldots,R_n\}$ be a family of attribute sets with $R_1 \succ R_2 \succ \cdots \succ R_n$ $(R_i \in 2^A, i = 1,2\ldots,n)$. The approximation precision $\alpha_P(F)$ is defined as

$$\alpha_P(F) = \alpha_P(F^-) \wedge \alpha_P(F^+) = \frac{\sum_{x\in U}\mu_{\underline{apr}_P(F^-)}(x)}{\sum_{x\in U}\mu_{\overline{apr}_P(F^-)}(x)} \wedge \frac{\sum_{x\in U}\mu_{\underline{apr}_P(F^+)}(x)}{\sum_{x\in U}\mu_{\overline{apr}_P(F^+)}(x)}$$

where $F \neq \emptyset$.

Theorem 2. *Let $S = (U,A)$ be an information system, F be an interval-valued fuzzy set of U and $P = \{R_1,R_2,\ldots,R_n\}$ be a family of attribute sets with $R_1 \succ R_2 \succ \cdots \succ R_n$ $(R_i \in 2^A, i = 1,2\ldots,n)$. Let $P_i = \{R_1,R_2,\ldots,R_i\}$, then for $\forall P_i$, $(i = 1,2,\ldots,n)$, we have*

$$\alpha_{P_1}(F) \leqslant \alpha_{P_2}(F) \leqslant \cdots \leqslant \alpha_{P_n}(F).$$

Proof. *For* $\forall x \in U$, $\mu_{\overline{apr}_{P_1}(F^-)}(x) = \sup_{x\in[x]_{R_1}}\mu_{F^-}(x) \geqslant \sup_{x\in[x]_{R_2}}\mu_{F^-}(x) = \mu_{\overline{apr}_{P_2}(F^-)}(x) \geqslant \cdots \geqslant \mu_{\overline{apr}_{P_n}(F^-)}(x)$. Moreover, it is clear from Theorem 1 that $\mu_{\underline{apr}_{P_1}(F^-)}(x) \leqslant \mu_{\underline{apr}_{P_2}(F^-)}(x) \leqslant \cdots \leqslant \mu_{\underline{apr}_{P_n}(F^-)}(x)$, thus according to Definition 8, we can easily obtain $\alpha_{P_1}(F^-) \leqslant \alpha_{P_2}(F^-) \leqslant \cdots \leqslant \alpha_{P_n}(F^-)$. Similarly, we can show $\alpha_{P_1}(F^+) \leqslant \alpha_{P_2}(F^+) \leqslant \cdots \leqslant \alpha_{P_n}(F^+)$. Therefore, $\alpha_{P_1}(F) \leqslant \alpha_{P_2}(F) \leqslant \cdots \leqslant \alpha_{P_n}(F)$.

The approximation precision is introduced to the forward approximation in order to describe the uncertainty of a target concept under a granulation order. Theorem 2 states that the approximation precision $\alpha_P(F)$ increases as the granulation order becomes longer. □

4.2. Fuzzy rule extraction algorithm based on forward approximation

In the rough fuzzy set theory, rule extraction is usually performed under uniform granulation, so the dynamic property is deficient in the process of rule extraction. However, we often need to extract rule dynamically according to the user's requirements in practical application. In a decision table, granulation is mainly reflected as the hierarchy relation between condition and decision attribute sets. The forward approximation approaches a target concept by the change in granulation, which can fully embody this hierarchy relation. The universe dwindles gradually and the approximation precision increases monotonously as the granulation order becomes longer. Thus the computational complexity of the algorithm is effectively reduced. Based on the forward approximation, a rule extraction algorithm called MRBFA is proposed.

Definition 9. Let $S = (U, C \cup D)$ be a decision table, where C is a set of condition attributes,D is a set of interval-valued fuzzy decision attributes. The positive region of D with regard to C is defined as:

$$\mu_{pos_C(D)}(x) = \sup_{F\in U/D}\mu_{\underline{apr}_C(F)}(x) = [\sup_{F\in U/D}\mu_{\underline{apr}_C(F^-)}(x), \sup_{F\in U/D}\mu_{\underline{apr}_C(F^+)}(x)].$$

The dependency degree $\gamma_C(D)$ of C with regard to D is defined as:

$$\gamma_C(D) = \frac{\sum_{x\in U}\sup_{F\in U/D}\mu_{\underline{apr}_C(F^-)}(x) + \sum_{x\in U}\sup_{F\in U/D}\mu_{\underline{apr}_C(F^+)}(x)}{2|U|}.$$

Based on the forward approximation, we propose a rule extraction algorithm.

4.2.1. Algorithm MRBFA (mining rules based on the forward approximation)

Input:decision table with interval-valued fuzzy decision attribute $S = (U, C \cup D)$
Output:decision rules.
(1) For $\forall c \in C$, compute the dependency degree $\gamma_{\{c\}}(D)$, let $\gamma_{\{c_1\}}(D) = \max\{\gamma_{\{c\}}(D) | c \in C\}$ and $P_1 = \{c_1\}$;
(2) $U/D = \{F_1, F_2, \ldots, F_d\}$, where, F_k $(k = 1, 2, \ldots, d)$ is an interval-valued fuzzy set;
(3) Let $P = \{P_1\}, U_1 = U, Rule' = Rule = \emptyset$, $i = 1$;
(4) Let $W_i = \cup_{k=1}^{d}\{x | \mu_{\underline{apr_P}(F_k^-)}(x) = \inf_{x \in U_i/P_i} \mu_{F_k^-}(x) \geqslant \eta^-, \mu_{\underline{apr_P}(F_k^+)}(x) = \inf_{x \in U_i/P_i} \mu_{F_k^+}(x) \geqslant \eta^+\}$. If $W_i \neq \emptyset$, then for $\forall x \in W_i$, put $des_P(x) \to des_{F_k}(x) (k = 1, 2, \ldots, d)$ into $Rule'$. Let $Rule = Rule \cup Rule'$ and $U_{i+1} = U_i - W_i$;
(5) If $C - P = \emptyset$ and $U_{i+1} \neq \emptyset$, then for $\forall x \in U_{i+1}$, let $T = \{x | \mu_{\underline{apr_P}(F)}(x) \geqslant_w \eta\} = \cup_{k=1}^{d}\{x | \mu_{\underline{apr_P}(F_k^-)}(x) + \mu_{\underline{apr_P}(F_k^+)}(x) \geqslant \eta^- + \eta^+, \mu_{\underline{apr_P}(F_k^+)}(x) - \mu_{\underline{apr_P}(F_k^-)}(x) \geqslant \eta^+ - \eta^-\}$. For $\forall x \in T$, put $des_P(x) \to des_{F_k}(x)$ $(k = 1, 2, \ldots, d)$ into $Rule$, go to (8);
(6) If $U_{i+1} = \emptyset$, go to (8);
(7) For $\forall c \in C - P$, compute $\gamma_{P \cup \{c\}}(D)$, let $\gamma_{P \cup \{c_2\}}(D) = \max\{\gamma_{P \cup \{c\}}(D) | c \in C - P\}$. Let $P_{i+1} = P_i \cup \{c_2\}$, $P = P \cup P_{i+1}$, $i = i + 1$, go to (4);
(8) Output $Rule$.

Remark 3. In Step (4), $\eta = [\eta^-, \eta^+]$ is a threshold and $\eta^-, \eta^+ \in [0.5, 1]$. In general, more conditions must be satisfied in the rules and the applicability of the rules decreases with the increase of η. That is, η determines the granulation of the fuzzy rules to some extent. The selection of η is determined by the actual requirement provided by the user.

Remark 4. In Step (4), $des_P(x)$ is the antecedent of the rule, and $des_{F_k}(x)$ is the consequent. For example, "If a_3 is 1 Then d is F_2", where, "a_3 is 1" is $des_P(x)$, "d is F_2" is $des_{F_k}(x)$.

According to MRBFA, one can extract a family of decision rules with granulations changing from coarse to fine. The dynamic classification results can approximate the decision classification as much as possible. MRBFA not only fully considers the potential community characters among objects, but also possesses high efficiency. The time complexity to extract rules is polynomial.

In Step (1), the time complexity for computing $\gamma_c(D)$ is $O(|C||U|^2)$.
In Step (2), the time complexity for computing U/D is $O(|U|^2)$.
In Step (4), the time complexity for computing W_i is $O(|P_i||U_i|^2)$.
In Step (5), the time complexity for computing T is $O(|C||U_{i+1}|^2)$.
In Step (7), the time complexity for computing $\gamma_{P \cup \{c\}}(D)$ is $O(|C - P_i|(|P_i| + 1)|U_{i+1}|^2)$.

From Steps (4) to (7), $|C|$ is the maximum value of the circle times. Therefore, the time complexity is

$$\sum_{i=1}^{|C|}(O(|P_i||U_i|^2) + O(|C||U_{i+1}|^2) + O(|C - P_i|(|P_i| + 1)|U_{i+1}|^2)). \qquad (*)$$

It is obvious that $|P_i| \leqslant |C|$, $|U_i| \leqslant |U|$, $|U_{i+1}| < |U|$, thus, the time complexity of $(*)$ is smaller than $O(|C|^3|U|^2)$. Other Steps will not be considered because that their time complexities are all constants. Thus the time complexity of the algorithm MABFA is

$$O(|C||U|^2) + O(|U|^2) + \sum_{i=1}^{|C|}(O(|P_i||U_i|^2) + O(|C||U_{i+1}|^2) + O(|C - P_i|(|P_i| + 1)|U_{i+1}|^2)) \leqslant O(|C|^3|U|^2).$$

Usually, the time complexity of the rule extraction algorithm based on attribute reduction is $O(|C|^3|U|^2)$, MRBFA is much smaller due to the universe dwindles gradually.

Remark 5. The main differences between MRBFA and the method based on attribute reduction include two aspects. First, rule extraction is based on a granulation order; thus, the adverse effects of attribute reduction are excluded as much as possible. Second, the time complexity of the model is effectively reduced because of the dwindling universe.

4.3. An example

To illustrate the operation of MRBFA, an example is given here. The decision table comes from Table 2. where $U = \{x_1, x_2, \ldots, x_{10}\}$ is a set of objects, $C = \{a_1, a_2, a_3\}$ is a condition attribute set, d is a decision attribute, separated into three linguistic terms F_1, F_2, F_3 and F_1, F_2, F_3 are interval-valued fuzzy sets.

According to MRBFA, compute the dependency degrees of a_1, a_2, a_3 with regard to d respectively. We can obtain $\gamma_{\{a_1\}}(d) = \frac{83}{200}$, $\gamma_{\{a_2\}}(d) = \frac{69}{200}$, $\gamma_{\{a_3\}}(d) = \frac{92}{200}$.

Hence, $P_1 = \{a_3\}$ and $P = \{P_1\}$, $U_1 = U$. For $U_1/a_3 = \{\{x_1, x_3, x_4, x_9\}, \{x_2, x_7, x_{10}\}, \{x_5, x_8\}, \{x_6\}\}$,

$$\text{when } x \in \{x_1, x_3, x_4, x_9\}: \mu_{\underline{apr_P}(F_1^-)}(x) = 0.15, \quad \mu_{\underline{apr_P}(F_1^+)}(x) = 0.2, \quad \mu_{\underline{apr_P}(F_2^-)}(x) = 0.15, \quad \mu_{\underline{apr_P}(F_2^+)}(x) = 0.2, \quad \mu_{\underline{apr_P}(F_3^-)}(x) = 0.1, \quad \mu_{\underline{apr_P}(F_3^+)}(x) = 0.2;$$

$$\text{when } x \in \{x_2, x_7, x_{10}\}: \mu_{\underline{apr_P}(F_1^-)}(x) = 0.05, \quad \mu_{\underline{apr_P}(F_1^+)}(x) = 0.1, \quad \mu_{\underline{apr_P}(F_2^-)}(x) = 0.5, \quad \mu_{\underline{apr_P}(F_2^+)}(x) = 0.7, \quad \mu_{\underline{apr_P}(F_3^-)}(x) = 0.05, \quad \mu_{\underline{apr_P}(F_3^+)}(x) = 0.2;$$

$$\text{when } x \in \{x_5, x_8\}: \mu_{\underline{apr_P}(F_1^-)}(x) = 0.05, \quad \mu_{\underline{apr_P}(F_1^+)}(x) = 0.1, \quad \mu_{\underline{apr_P}(F_2^-)}(x) = 0.2, \quad \mu_{\underline{apr_P}(F_2^+)}(x) = 0.3, \quad \mu_{\underline{apr_P}(F_3^-)}(x) = 0.5, \quad \mu_{\underline{apr_P}(F_3^+)}(x) = 0.6;$$

$$\text{when } x \in \{x_6\}: \mu_{\underline{apr_P}(F_1^-)}(x) = 0.1, \quad \mu_{\underline{apr_P}(F_1^+)}(x) = 0.2, \quad \mu_{\underline{apr_P}(F_2^-)}(x) = 0.35, \quad \mu_{\underline{apr_P}(F_2^+)}(x) = 0.5, \quad \mu_{\underline{apr_P}(F_3^-)}(x) = 1.0, \quad \mu_{\underline{apr_P}(F_3^+)}(x) = 1.0.$$

Let $\eta = [\eta^-, \eta^+] = [0.5, 0.6]$. Notice that for $x \in \{x_2, x_7, x_{10}\}$, $\mu_{\underline{apr_P}(F_2^-)}(x) = 0.5 \geqslant \eta^-$, $\mu_{\underline{apr_P}(F_2^+)}(x) = 0.7 \geqslant \eta^+$; $x \in \{x_5, x_8\}$, $\mu_{\underline{apr_P}(F_3^-)}(x) = 0.5 \geqslant \eta^-$, $\mu_{\underline{apr_P}(F_3^+)}(x) = 0.6 \geqslant \eta^+$; $x \in \{x_6\}$, $\mu_{\underline{apr_P}(F_3^-)}(x) = 1.0 \geqslant \eta^-$, $\mu_{\underline{apr_P}(F_3^+)}(x) = 1.0 \geqslant \eta^+$, then we have

$$W_1 = \{x_2, x_5, x_6, x_7, x_8, x_{10}\},$$
$$Rule = \{r_1 : des_{\{a_3\}}(x_2, x_7, x_{10}) \to des_{F_2}(x);$$
$$r_2 : des_{\{a_3\}}(x_5, x_8) \to des_{F_3}(x);$$
$$r_3 : des_{\{a_3\}}(x_6) \to des_{F_3}(x)\} \text{ and } U_2 = U_1 - W_1 = \{x_1, x_3, x_4, x_9\}.$$

For $C - P \neq \emptyset$ and $U_2 \neq \emptyset$, continue to compute the dependency degrees of the rest attributes a_1, a_2 with respect to D. It should be noted that the computation of the dependency degree is based on the updated U_2, we obtain $\gamma_{\{a_1,a_3\}}(D) = \frac{140}{800}, \gamma_{\{a_2,a_3\}}(D) = \frac{475}{800}$. So choose a_2 as c_2, $P_2 = \{a_2, a_3\}$ and $P = \{a_2, a_3\}$. $U_2/P = \{\{x_1, x_3, x_9\}, \{x_4\}\}$.

$$\text{when } x \in \{x_1, x_3, x_9\}: \mu_{\underline{apr_P}(F_1^-)}(x) = 0.45, \quad \mu_{\underline{apr_P}(F_1^+)}(x) = 0.7, \quad \mu_{\underline{apr_P}(F_2^-)}(x) = 0.15, \quad \mu_{\underline{apr_P}(F_2^+)}(x) = 0.2, \quad \mu_{\underline{apr_P}(F_3^-)}(x) = 0.1, \quad \mu_{\underline{apr_P}(F_3^+)}(x) = 0.2;$$

when $x \in \{x_4\}: \mu_{\underline{apr}_P(F_1^-)}(x) = 0.15, \quad \mu_{\underline{apr}_P(F_1^+)}(x) = 0.2,$ $\mu_{\underline{apr}_P(F_2^-)}(x) = 0.5, \quad \mu_{\underline{apr}_P(F_2^+)}(x) = 0.8, \quad \mu_{\underline{apr}_P(F_3^-)}(x) = 0.2,$ $\mu_{\underline{apr}_P(F_3^+)}(x) = 0.3;$

Notice that for $x \in \{x_4\}, \mu_{\underline{apr}_P(F_2^-)}(x) = 0.5 \geqslant \eta^-, \mu_{\underline{apr}_P(F_2^+)}(x) = 0.8 \geqslant \eta^+$, then $W_2 = \{x_4\}$ and $Rule = \{r_1 : des_{\{a_3\}}(x_2, x_7, x_{10}) \to des_{F_2}(x);\ r_2 : des_{\{a_3\}}(x_5, x_8) \to des_{F_3}(x); r_3 : des_{\{a_3\}}(x_6) \to des_{F_3}(x);\ r_4 : des_{\{a_2,a_3\}}(x_4) \to des_{F_2}(x)\}$ and $U_3 = U_2 - W_2 = \{x_1, x_3, x_9\}$. For $C - P \neq \emptyset$ and $U_3 \neq \emptyset$, the last attribute a_1 is added to P, that is $P = \{a_1, a_2, a_3\}$. Then $U_3 /P = \{\{x_1, x_3, x_9\}\}$.

when $x \in \{x_1, x_3, x_9\}$: $\mu_{\underline{apr}_P(F_1^-)}(x) = 0.45, \mu_{\underline{apr}_P(F_1^+)}(x) = 0.7, \mu_{\underline{apr}_P(F_2^-)}(x) = 0.15, \mu_{\underline{apr}_P(F_2^+)}(x) = 0.2, \mu_{\underline{apr}_P(F_3^-)}(x) = 0.1, \mu_{\underline{apr}_P(F_3^+)}(x) = 0.2.$

It is clear that $W_3 = \emptyset$ and $U_4 = U_3$. Because of $C - P = \emptyset$ and $U_4 \neq \emptyset$, for $x \in U_4 = \{x_1, x_3, x_9\}$, $\mu_{\underline{apr}_P(F_1^-)}(x) + \mu_{\underline{apr}_P(F_1^+)}(x) = 0.45 + 0.7 \geqslant \eta^- + \eta^+ = 0.5 + 0.6$, $\mu_{\underline{apr}_P(F_1^+)}(x) - \mu_{\underline{apr}_P(F_1^-)}(x) = 0.7 - 0.45 \geqslant \eta^+ - \eta^- = 0.6 - 0.5$, that is $\mu_{\underline{apr}_P(F_1)}(x) \geqslant_w \eta$, then $T = \{x_1, x_3, x_9\}$ and $r_5 : des_{\{a_1,a_2,a_3\}}(x_1, x_3, x_9) \to des_{F_1}(x)$ is added to $Rule$. The algorithm stopped and rules are obtained as follows:

$Rule = \{r_1 : des_{\{a_3\}}(x_2, x_7, x_{10}) \to des_{F_2}(x)$, namely: If a_3 is 1 Then d is F_2 and $\mu_{\underline{apr}_P(F_2)}(x) \geqslant [0.5, 0.7]$; $r_2 : des_{\{a_3\}}(x_5, x_8) \to des_{F_3}(x)$, namely: If a_3 is 4 Then d is F_3 and $\mu_{\underline{apr}_P(F_3)}(x) \geqslant [0.5, 0.6]$; $r_3 : des_{\{a_3\}}(x_6) \to des_{F_3}(x)$, namely: If a_3 is 2 Then d is F_3 and $\mu_{\underline{apr}_P(F_3)}(x) = [1.0, 1.0]$; $r_4 : des_{\{a_2,a_3\}}(x_4) \to des_{F_2}(x)$, namely: If a_2 is 2 and a_3 is 3. Then d is F_2 and $\mu_{\underline{apr}_P(F_2)}(x) \geqslant [0.5, 0.8]$; $r_5 : des_{\{a_1,a_2,a_3\}}(x_1, x_3, x_9) \to des_{F_1}(x)$, namely: If a_1 is 2, a_2 is 1 and a_3 is 3 Then d is F_1 and $\mu_{\underline{apr}_P(F_1)}(x) \geqslant [0.45, 0.7]\}$.

5. Backward approximation in interval-valued rough fuzzy sets

The main purpose of the forward approximation is extending the interval-valued rough fuzzy sets approximation from static granulation to dynamic one, and approaching a target concept by the change of granulation. Through the forward approximation, one can extract a family of fuzzy decision rules with granulations changing from coarse to fine. In some applications, however, the approximation precision is restricted by the decision requirements or preference of decision makers (Qian, 2008). An obvious problem is extracting simpler rules based on keeping the approximation precision invariant. The forward approximation appears unsuitable for this purpose. Therefore, the backward approximation in interval-valued rough fuzzy sets is proposed.

In the process of the backward approximation, the objects needing to be further investigated in the universe are considered as the next researched objective. Then a sequence of expressions with different granulation levels can be generated. In the family of equivalence relations, the backward approximation can not only effectively reduce the knowledge granules describing the interval-valued fuzzy set, it also fully mines potential community characteristics among objects based on keeping the approximation precision invariant.

5.1. The concept of backward approximation

Let $S = (U, A)$ be an information system, F be an interval-valued fuzzy set of U and $P = \{R_1, R_2, \ldots, R_n\}$ be a family of attribute sets with$R_1 \prec R_2 \prec \cdots \prec R_n$ ($R_i \in 2^A$, $i = 1, 2 \ldots, n$). So the sequence of attribute sets $R_i \in 2^A$ ($i = 1, 2, \ldots, n$) can determine a sequence of granulation spaces from fine to coarse. The upper and lower approximations of backward approximation are defined under a granulation order.

Definition 10. Let $S = (U,A)$ be an information system, F be an interval-valued fuzzy set of U and $P = \{R_1, R_2, \ldots, R_n\}$ be a family of attribute sets with$R_1 \prec R_2 \prec \cdots \prec R_n (R_i \in 2^A, i = 1, 2 \ldots, n)$. Let $P_i = \{R_1, R_2, \ldots, R_i\}$, P_i-upper approximation $\overline{apr}_{P_i}(F)$ and P_i-lower approximation $\underline{apr}_{P_i}(F)$ of P_i-backward approximation of F are defined as

$$\mu_{\overline{apr}_{P_i}(F)}(x) = [\mu_{\overline{apr}_{P_i}(F^-)}(x), \mu_{\overline{apr}_{P_i}(F^+)}(x)] = [\sup_{x \in [x]_{R_1}} \mu_{F^-}(x), \sup_{x \in [x]_{R_1}} \mu_{F^+}(x)],$$

$$\mu_{\underline{apr}_{P_i}(F)}(x) = [\mu_{\underline{apr}_{P_i}(F^-)}(x), \mu_{\underline{apr}_{P_i}(F^+)}(x)] = \begin{cases} [\vee_{h=1}^{j} \inf_{x \in [x]_{R_h}} \mu_{F^-}(x), \vee_{h=1}^{j} \inf_{x \in [x]_{R_h}} \mu_{F^+}(x)], & \text{if } \exists j \\ [\inf_{x \in [x]_{R_1}} \mu_{F^-}(x), \inf_{x \in [x]_{R_1}} \mu_{F^+}(x)], & \text{otherwise} \end{cases}$$

where, $j = \max\{t | \mu_{\underline{apr}_{R_t}(F)}(x) = [\inf_{x \in [x]_{R_t}} \mu_{F^-}(x), \inf_{x \in [x]_{R_t}} \mu_{F^+}(x)] \geqslant [\zeta^-, \zeta^+], 1 \leqslant t \leqslant i\}$, ζ^-, $\zeta^+ \in [0.5, 1]$ and $\zeta = [\zeta^-, \zeta^+] \in [I]$ is a suitable threshold.

The boundary $BN_P(F)$ of F is defined as:

$$\mu_{BN_{P_i}(F)}(x) = [\sup_{y \in [x]_{\cup_{R_i \in P_i} R_i}} \mu_{F^-}(y) \wedge (1 - \inf_{y \in [x]_{\cup_{R_i \in P_i} R_i}} \mu_{F^+}(y)), \sup_{y \in [x]_{\cup_{R_i \in P_i} R_i}} \mu_{F^+}(y) \wedge (1 - \inf_{y \in [x]_{\cup_{R_i \in P_i} R_i}} \mu_{F^-}(y))].$$

Remark 6. The backward approximation is concentrated on the change in the construction of the target concept. The main idea of Definition 10 is that the number of equivalence classes used to describe the target concept reduced as the granulation order becomes longer. That is, new equivalence classes under different granulations are induced by combining known equivalence classes. Although the upper and lower approximation operators are equivalent to the ones in Definition 4, the backward approximation emphasizes the change in the construction of the target concept. The structure of the approximation operators reflects the granulation spaces changing from fine to coarse.

Remark 7. Notice that the upper and lower approximations are not symmetrical. For the sake of conciseness, the upper approximation operator is not denoted in structural form. Moreover, computing the upper approximation is not always necessary in many applications. In MRBBA (Section 5.2) only the lower approximation operator is used. Of course, one can also represent the upper approximation operator same as the structure of the lower one.

Remark 8. $\zeta = [\zeta^-, \zeta^+]$ is a threshold and $\zeta^-, \zeta^+ \in [0.5, 1]$. In general, more conditions must be satisfied in the rules, and the applicability of the rules decreases with increasing ζ. That is, ζ ascertains the granulation of the fuzzy rules to some extent. The selection of ζ is determined by actual requirement provided by user.

Definition 10 shows that a target concept can be approached by the upper approximation $\overline{apr}_{P_i}(F)$ and the variable lower approximation $\underline{apr}_{P_i}(F)$. In particular, when $i = n$, we denote $\overline{apr}_{P_n}(F)$ as $\overline{apr}_P(F)$ and $\underline{apr}_{P_n}(F)$ as $\underline{apr}_P(F)$. $\overline{apr}_P(F)$ and $\underline{apr}_P(F)$ are called P-upper approximation and P-lower approximation of P-backward approximation of F, respectively.

Theorem 3. *Let $S = (U,A)$ be an information system, F be an interval-valued fuzzy set of U and $P = \{R_1, R_2, \ldots, R_n\}$ be a family of attribute sets with $R_1 \prec R_2 \prec \cdots \prec R_n (R_i \in 2^A, i = 1, 2 \ldots, n)$. Let $P_i = \{R_1, R_2, \ldots, R_i\}$, then for $\forall P_i$, $(i = 1, 2, \ldots, n)$, the following properties hold:*

$$\underline{apr}_{P_1}(F) = \underline{apr}_{P_2}(F) = \cdots = \underline{apr}_{P_n}(F), \tag{4}$$

$$\underline{apr}_{P_i}(F) \subseteq F \subseteq \overline{apr}_{P_i}(F), \tag{5}$$

$$BN_{P_1}(F) = BN_{P_2}(F) = \cdots = BN_{P_n}(F). \quad (6)$$

Proof. For $\forall P_i$, if there exists a j, then $\mu_{\underline{apr}_{P_i}(F)}(x) = [\vee_{h=1}^{j}\inf_{x\in[x]_{R_h}} \mu_{F^-}(x), \vee_{h=1}^{j}\inf_{x\in[x]_{R_h}} \mu_{F^+}(x)] = [\inf_{x\in[x]_{R_1}} \mu_{F^-}(x), \inf_{x\in[x]_{R_1}} \mu_{F^+}(x)] = \mu_{\underline{apr}_{R_1}(F)}(x)$; if there does not exist a j, then $\mu_{\underline{apr}_{P_i}(F)}(x) = [\inf_{x\in[x]_{R_1}} \mu_{F^-}(x), \inf_{x\in[x]_{R_1}} \mu_{F^+}(x)] = \mu_{\underline{apr}_{R_1}(F)}(x)$. Therefore, one can obtain $\underline{apr}_{P_1}(F) = \underline{apr}_{P_2}(F) = \cdots = \underline{apr}_{P_n}(F) = \underline{apr}_{R_1}(F)$, that is (4).

Moreover, $\mu_{\underline{apr}_{P_i}(F)}(x) = \mu_{\underline{apr}_{R_1}(F)}(x) = [\inf_{x\in[x]_{R_1}} \mu_{F^-}(x), \inf_{x\in[x]_{R_1}} \mu_{F^+}(x)] \leqslant [\mu_{F^-}(x), \mu_{F^+}(x)] = \mu_F(x) \leqslant [\sup_{x\in[x]_{R_1}} \mu_{F^-}(x), \sup_{x\in[x]_{R_1}} \mu_{F^+}(x)] = \mu_{\overline{apr}_{P_i}(F)}(x)$. Thus, $\underline{apr}_{P_i}(F) \subseteq F \subseteq \overline{apr}_{P_i}(F)$.

For $\forall i$, $\mu_{BN_{P_i}(F)}(x) = [\sup_{y\in[x]_{\cup_{R_i\in P_i}R_i}} \mu_{F^-}(y) \wedge (1-\inf_{y\in[x]_{\cup_{R_i\in P_i}R_i}} \mu_{F^+}(y)), \sup_{y\in[x]_{\cup_{R_i\in P_i}R_i}} \mu_{F^+}(y) \wedge (1-\inf_{y\in[x]_{\cup_{R_i\in P_i}R_i}} \mu_{F^-}(y))] = [\sup_{y\in[x]_{R_1}} \mu_{F^-}(y) \wedge (1-\inf_{y\in[x]_{R_1}} \mu_{F^+}(y)), \sup_{y\in[x]_{R_1}} \mu_{F^+}(y) \wedge (1-\inf_{y\in[x]_{R_1}} \mu_{F^-}(y)] = \mu_{BN_{R_1}(F)}(x)$. Then (6) is proved.

This completes the proof. □

Theorem 3 states that the lower and upper approximations of P-backward approximation remain invariant as the granulation order becomes longer.

Theorem 4. *Let S = (U,A) be an information system, F be an interval-valued fuzzy set of U and P = {$R_1, R_2, \ldots, R_n$} be a family of attribute sets with $R_1 \prec R_2 \prec \cdots \prec R_n (R_i \in 2^A, i = 1,2\ldots,n)$. Let P_i = {$R_1, R_2, \ldots, R_i$}, then for $\forall P_i$, (i = 1,2,...,n), the following properties hold:*

$$\alpha_{P_1}(F) = \alpha_{P_2}(F) = \cdots = \alpha_{P_n}(F),$$

where $\alpha_{P_i}(F) = \alpha_{P_i}(F^-) \wedge \alpha_{P_i}(F^+) = \frac{\sum_{x\in U}\mu_{\underline{apr}_{P_i}(F^-)}(x)}{\sum_{x\in U}\mu_{\overline{apr}_{P_i}(F^-)}(x)} \wedge \frac{\sum_{x\in U}\mu_{\underline{apr}_{P_i}(F^+)}(x)}{\sum_{x\in U}\mu_{\overline{apr}_{P_i}(F^+)}(x)}$ *is approximation precision.*

Proof. It follows from Definition 10 that $\overline{apr}_{P_1}(F) = \overline{apr}_{P_2}(F) = \cdots = \overline{apr}_{P_n}(F) = \overline{apr}_{R_1}(F)$. Then for $\forall x \in U$, $\mu_{\overline{apr}_{P_1}(F^-)}(x) = \mu_{\overline{apr}_{P_2}(F^-)}(x) = \cdots = \mu_{\overline{apr}_{P_n}(F^-)}(x)$, $\mu_{\overline{apr}_{P_1}(F^+)}(x) = \mu_{\overline{apr}_{P_2}(F^+)}(x) = \cdots = \mu_{\overline{apr}_{P_n}(F^+)}(x)$. And from Theorem 3, we can obtain that $\underline{apr}_{P_1}(F) = \underline{apr}_{P_2}(F) = \cdots = \underline{apr}_{P_n}(F)$, then for $\forall x \in U$, $\mu_{\underline{apr}_{P_1}(F^-)}(x) = \mu_{\underline{apr}_{P_2}(F^-)}(x) = \cdots = \mu_{\underline{apr}_{P_n}(F^-)}(x)$, $\mu_{\underline{apr}_{P_1}(F^+)}(x) = \mu_{\underline{apr}_{P_2}(F^+)}(x) = \cdots = \mu_{\underline{apr}_{P_n}(F^+)}(x)$. Therefore, $\alpha_{P_1}(F) = \alpha_{P_2}(F) = \cdots = \alpha_{P_n}(F)$. This completes the proof. □

Definition 11. Let $S = (U,A)$ be an information system, $R \in 2^A$ be a subset of attributes of$A$, $\Gamma = \{F_1, F_2, \ldots, F_m\}$ be a fuzzy partition of U, whereF_k $(k = 1,2,\ldots,m)$ is an interval-valued fuzzy set. Lower and upper approximations of Γ with respect to R are defined by

$$\underline{apr}_R\Gamma = \left\{\underline{apr}_R(F_1), \underline{apr}_R(F_2), \ldots, \underline{apr}_R(F_m)\right\},$$
$$\overline{apr}_R\Gamma = \{\overline{apr}_R(F_1), \overline{apr}_R(F_2), \ldots, \overline{apr}_R(F_m)\}.$$

The target fuzzy concept is described by equivalence classes. For the given universe U, the equivalence classes are determined by R. So the lower and upper approximations of Γ have a close relationship with R. We need to define a new measure to evaluate the convergence of $\Gamma$with respect to R, which is helpful in understanding the construction of the lower approximation.

Definition 12. Let $S = (U,A)$ be an information system, $R \in 2^A$ be a subset of attributes of A and $\Gamma = \{F_1,F_2,\ldots,F_m\}$ be a fuzzy partition of U, whereF_k $(k = 1,2,\ldots,m)$ is an interval-valued fuzzy set. Convergence degree of Γ with respect to R is defined as:

$$C(R,\Gamma) = C(R,\Gamma^-) \wedge C(R,\Gamma^+) = \left(\sum_{k=1}^{m}\frac{|F_k^-|}{|U|}\sum_{j=1}^{s_k}p^2(F_k^{-j})\right) \wedge \left(\sum_{k=1}^{m}\frac{|F_k^+|}{|U|}\sum_{j=1}^{s_k}p^2(F_k^{+j})\right),$$

where, $|F_k^-| = \sum_{x\in U}\mu_{F_k^-}(x)$, $|F_k^+| = \sum_{x\in U}\mu_{F_k^+}(x)$, $p(F_k^{-j}) = \frac{|F_k^{-j}|}{|F_k^-|}$, $p(F_k^{+j}) = \frac{|F_k^{+j}|}{|F_k^+|}$. s_k is the number of equivalence classes that satisfied $\mu_{\underline{apr}_R(F_k)}(x) = [\inf_{x\in[x]_R}\mu_{F_k^-}(x), \inf_{x\in[x]_R}\mu_{F_k^+}(x)] \geqslant [\zeta^-, \zeta^+] = \zeta$. Let $M_{Rk} = \{[x]_R | \mu_{\underline{apr}_R(F_k)}(x) = [\inf_{x\in[x]_R}\mu_{F_k^-}(x), \inf_{x\in[x]_R}\mu_{F_k^+}(x)] \geqslant [\zeta^-, \zeta^+] = \zeta,\ x \in U\}$, then $s_k = |M_{Rk}|$.

Remark 9. Without loss of generality, one can assume that $M_{Rk} = \{A_1, A_2, \ldots, A_{s_k}\}$, where, $A_j = [x_{t_j}]_R$, $j = 1,2,\ldots,s_k, t_j \in \{1,2,\ldots,|U|\}$ then convergence degree is denoted as

$$C(R,\Gamma) = C(R,\Gamma^-) \wedge C(R,\Gamma^+) = \left(\sum_{k=1}^{m}\frac{\sum_{x\in U}\mu_{F_k^-}(x)}{|U|}\sum_{j=1}^{s_k}\left(\frac{\sum_{x\in A_j}\mu_{F_k^-}(x)}{\sum_{x\in U}\mu_{F_k^-}(x)}\right)^2\right) \wedge \left(\sum_{k=1}^{m}\frac{\sum_{x\in U}\mu_{F_k^+}(x)}{|U|}\sum_{j=1}^{s_k}\left(\frac{\sum_{x\in A_j}\mu_{F_k^+}(x)}{\sum_{x\in U}\mu_{F_k^+}(x)}\right)^2\right).$$

Remark 10. If Γ = {F}, then

$$C(R,\Gamma) = C(R,\Gamma^-) \wedge C(R,\Gamma^+) = \left(\sum_{j=1}^{s}p^2(F^{-j})\right) \wedge \left(\sum_{j=1}^{s}p^2(F^{+j})\right) = \left(\frac{1}{|F^-|^2}\sum_{j=1}^{s}\left(\sum_{x\in[x_{t_j}]_R}\mu_{F^-}(x)\right)^2\right) \wedge \left(\frac{1}{|F^+|^2}\sum_{j=1}^{s}\left(\sum_{x\in[x_{t_j}]_R}\mu_{F^+}(x)\right)^2\right).$$

Definition 13. Let $S = (U,A)$ be an information system, $\Gamma = \{F_1,F_2,\ldots,F_m\}$ be a fuzzy partition of U, where F_k $(k = 1,2,\ldots,m)$ is an interval-valued fuzzy set. $P = \{R_1,R_2,\ldots,R_n\}$ be a family of attribute sets with $R_1 \prec R_2 \prec \cdots \prec R_n$ $(R_i \in 2^A, i = 1,2\ldots,n)$. Let $P_i = \{R_1,R_2,\ldots,R_i\}$, convergence degree of Γ with respect to P is defined as:

$$C(P,\Gamma) = C(P,\Gamma^-) \wedge C(P,\Gamma^+) = \left(\sum_{k=1}^{m}\frac{|F_k^-|}{|U|}\sum_{j=1}^{s_k}p^2(F_k^{-j})\right) \wedge \left(\sum_{k=1}^{m}\frac{|F_k^+|}{|U|}\sum_{j=1}^{s_k}p^2(F_k^{+j})\right),$$

where, $|F_k^-| = \sum_{x\in U}\mu_{F_k^-}(x)$, $|F_k^+| = \sum_{x\in U}\mu_{F_k^+}(x)$, $p(F_k^{-j}) = \frac{|F_k^{-j}|}{|F_k^-|}$, $p(F_k^{+j}) = \frac{|F_k^{+j}|}{|F_k^+|}$, $s_k = -M_{Pk}-$, $M_{Pk} = \{\bigcup_{h=1}^{j_x}[x]_{R_h} | j_x = \max\{t | \mu_{\underline{apr}_{R_t}(F_k)}(x) = [\inf_{x\in[x]_{R_t}}\mu_{F_k^-}(x), \inf_{x\in[x]_{R_t}}\mu_{F_k^+}(x)] \geqslant [\zeta^-, \zeta^+]\}\}$.

Remark 11. Without loss of generality, one can assume that $M_{Pk} = \{A_1, A_2, \ldots, A_{s_k}\}$, where, $A_j = [x_{t_j}]_{R_{I_j}}$, $j = 1,2,\ldots,s_k, I_j \in \{1,2,\ldots,n\}$, $t_j \in \{1,2,\ldots,|U|\}$, then convergence degree is represented as

$$C(P,\Gamma) = C(P,\Gamma^-) \wedge C(P,\Gamma^+) = \left(\sum_{k=1}^{m}\frac{\sum_{x\in U}\mu_{F_k^-}(x)}{|U|}\sum_{j=1}^{s_k}\left(\frac{\sum_{x\in A_j}\mu_{F_k^-}(x)}{\sum_{x\in U}\mu_{F_k^-}(x)}\right)^2\right) \wedge \left(\sum_{k=1}^{m}\frac{\sum_{x\in U}\mu_{F_k^+}(x)}{|U|}\sum_{j=1}^{s_k}\left(\frac{\sum_{x\in A_j}\mu_{F_k^+}(x)}{\sum_{x\in U}\mu_{F_k^+}(x)}\right)^2\right).$$

Remark 12. If $\Gamma=\{F\}$, then $C(P,\Gamma)=C(P,\Gamma^-)\wedge C(P,\Gamma^+)=(\sum_{j=1}^{s}p^2(F^{-j}))\wedge\left(\sum_{j=1}^{s}p^2(F^{+j})\right)=\left(\frac{1}{|F^-|^2}\sum_{j=1}^{s}(\sum_{x\in A_j}\mu_{F^-}(x))^2\right)\wedge\left(\frac{1}{|F^+|^2}\sum_{j=1}^{s}(\sum_{x\in A_j}\mu_{F^+}(x))^2\right)$, where $A_j=[x_{t_j}]_{R_{l_j}}$.

Theorem 5. *Let $S=(U,A)$ be an information system, F be an interval-valued fuzzy set of U and $P=\{R_1,R_2,\ldots,R_n\}$ be a family of attribute sets with $R_1\prec R_2\prec\cdots\prec R_n(R_i\in 2^A,\ i=1,2\ldots,n)$. Let $P_i=\{R_1,R_2,\ldots,R_i\}$, then for $\forall P_i$ $(i=1,2,\ldots,n)$, the following property holds:*

$$C(P_1,F)\leqslant C(P_2,F)\leqslant\cdots\leqslant C(P_n,F).$$

Proof. Suppose $1\leqslant a<b\leqslant n$, $M_{P_ak}=\{A_1,A_2,\ldots,A_m\}$, $M_{P_bk}=\{B_1,B_2,\ldots,B_n\}$. Obviously, $M_{P_ak}\subseteq M_{P_bk}$ and $m>n$. Then there may exist a partition $\{C_1,C_2,\ldots,C_n\}$ of $\{1,2,\ldots,m\}$ such that $B_t=\bigcup_{l\in C_t}A_l$, $t=1,2,\ldots,n$. Therefore, one can obtain $C(P_b,F)=C(P_b,F^-)\wedge C(P_b,F^+)=\left(\frac{1}{|F^-|^2}\sum_{t=1}^{s_b}(\sum_{x\in B_t}\mu_{F^-}(x))^2\right)\wedge\left(\frac{1}{|F^+|^2}\sum_{t=1}^{s_b}(\sum_{x\in B_t}\mu_{F^+}(x))^2\right)=\left(\frac{1}{|F^-|^2}\sum_{t=1}^{s_b}(\sum_{x\in\bigcup_{l\in C_t}A_l}\mu_{F^-}(x))^2\right)\wedge\left(\frac{1}{|F^+|^2}\sum_{t=1}^{s_b}(\sum_{x\in\bigcup_{l\in C_t}A_l}\mu_{F^+}(x))^2\right)=\left(\frac{1}{|F^-|^2}\sum_{t=1}^{s_b}(\sum_{l\in C_t}\sum_{x\in A_l}\mu_{F^-}(x))^2\right)\wedge\left(\frac{1}{|F^+|^2}\sum_{t=1}^{s_b}(\sum_{l\in C_t}\sum_{x\in A_l}\mu_{F^+}(x))^2\right)\geqslant\left(\frac{1}{|F^-|^2}\sum_{t=1}^{s_b}\sum_{l\in C_t}(\sum_{x\in A_l}\mu_{F^-}(x))^2\right)\wedge\left(\frac{1}{|F^+|^2}\sum_{t=1}^{s_b}\sum_{l\in C_t}(\sum_{x\in A_l}\mu_{F^+}(x))^2\right))=\left(\frac{1}{|F^-|^2}\sum_{l=1}^{s_a}(\sum_{x\in A_l}\mu_{F^-}(x))^2\right)\wedge\left(\frac{1}{|F^+|^2}\sum_{l=1}^{s_a}(\sum_{x\in A_l}\mu_{F^+}(x))^2\right)=C(P_a,F^-)\wedge C(P_a,F^+)=C(P_a,F)$.

Thus $C(P_1,F)\leqslant C(P_2,F)\leqslant\cdots\leqslant C(P_n,F)$. □

Theorem 6. *Let $S=(U,A)$ be an information system, $\Gamma=\{F_1,F_2,\cdots,F_m\}$ be a fuzzy partition of U, where F_k $(k=1,2,\cdots,m)$ is an interval-valued fuzzy set. $P=\{R_1,R_2,\ldots,R_n\}$ be a family of attribute sets with $R_1\prec R_2\prec\cdots\prec R_n(R_i\in 2^A,\ i=1,2\ldots,n)$. Let $P_i=\{R_1,R_2,\ldots,R_i\}$, then for $\forall P_i$, $(i=1,2,\ldots,n)$, the following property holds:*

$$C(P_1,\Gamma)\leqslant C(P_2,\Gamma)\leqslant\cdots\leqslant C(P_n,\Gamma).$$

Proof. It follows from Theorem 5 that $C(P_1,\ F_k)\leqslant C(P_2,\ F_k)\leqslant\cdots\leqslant C(P_n,\ F_k)$ for $\forall F_k$ $(k\leqslant m)$. Suppose $1\leqslant\ a<b\leqslant n$, then $C(P_a,\ F_k)\leqslant C(P_b,\ F_k)$. Therefore, we can obtain that $C(P_a,\Gamma)=(\sum_{k=1}^{m}\frac{|F_k^-|}{|U|}\sum_{j=1}^{s_k}p^2(F_k^{-j}))\wedge(\sum_{k=1}^{m}\frac{|F_k^+|}{|U|}\sum_{j=1}^{s_k}p^2(F_k^{+j}))=(\sum_{k=1}^{m}\frac{|F_k^-|}{|U|}\cdot C(P_a,F_k^-))\wedge\ (\sum_{k=1}^{m}\frac{|F_k^+|}{|U|}\cdot C(P_a,F_k^+))\leqslant(\sum_{k=1}^{m}\frac{|F_k^-|}{|U|}\cdot C(P_b,F_k^-))\wedge(\sum_{k=1}^{m}\frac{|F_k^+|}{|U|}\cdot C(P_b,F_k^+))=C(P_b,\Gamma)$. Thus $C(P_1,\Gamma)\leqslant C(P_2,\Gamma)\leqslant\cdots\leqslant C(P_n,\Gamma)$. □

Theorem 6 shows that the convergence degree of Γ with respect to P_i increases as the granulation order becomes longer. New equivalence classes under different granulations are induced by combining some known equivalence classes. Thus the number of equivalence classes for describing the target concept is reduced. This suggests a new idea to describe a target concept with as few equivalence classes as possible based on keeping the approximation precision invariant. This may have potential applications in the interval-valued rough fuzzy set theory, such as the description of multi-target concepts, approximation classification, and rule extraction.

5.2. Fuzzy rule extraction algorithm based on backward approximation

Interval-valued rough fuzzy sets theory is used to mine some fuzzy decision rules in the form of "if …, then …" from an interval-valued fuzzy decision tables. More exactly, the fuzzy decision rules say that if condition attributes have some given values, then decision attributes have the other given values. In this section, as an application of the backward approximation, we apply this approach to fuzzy-decision-rule extraction. In view of the backward approximation is based on dynamic granulation, the decision classification induced by decision attributes can be regarded as target concepts, and the condition attribute sets can be used to construct a granulation order. Based on the backward approximation, a rule extraction algorithm called MRBBA is designed.

Let $S=(U,C\cup D)$ be a decision table, for $\forall c\in C$, the significance of c with respect to D is defined as: $sig_C^D(c)=\gamma_C(D)-\gamma_{C-\{c\}}(D)$, where $\gamma_C(D)=\frac{\sum_{x\in U}\sup_{F\in U/D}\mu_{\underline{apr}_C(F^-)}(x)+\sum_{x\in U}\sup_{F\in U/D}\mu_{\underline{apr}_C(F^+)}(x)}{2|U|}$.

5.2.1. Algorithm MRBBA (mining rules based on the backward approximation)

Input: decision table $S=(U,C\cup D)$
Output: decision rules.

(1) Compute decision classes $U/D=\{F_1,F_2,\ldots,F_d\}$, where, F_k $(k=1,2,\ldots,d)$ is an interval-valued fuzzy set;
(2) Let $Rule=\emptyset$, $P_1=\{\{C_1\}\}$, $j=1$, $C_1=C$;
(3) For $\forall c\in C_j$, compute the significance $sig_{C_j}^D(c)$. Let $B=\{c_0|sig_{C_j}^D(c_0)=\min\{sig_{C_j}^D(c),c\in C_j\}\}$. If $|B|\neq 1$, then let $\gamma_{\{c_0\}}(D)=\min\{\gamma_{\{c\}}(D),c\in B\}$;
(4) Let $C_{j+1}=C_j-\{c_0\}$, $P_{j+1}=P_j\cup\{\{C_{j+1}\}\}$;
(5) $j=j+1$. If $j<|C|$, go to (3); Otherwise, go to (6);
(6) Let $k=1$;
(7) Let $P=P_j$. Compute $\underline{apr}_P(F_k)$, M_{Pk};
(8) Put every decision rule $des([x])\to des_{F_k}(x)$ into $Rule$, where $[x]\in M_{Pk}$;
(9) $k=k+1$. If $k\leqslant d$, go to (7); otherwise, go to (10);
(10) let $T=\bigcup_{k=1}^{d}\{x|\mu_{\underline{apr}_P(F_k)}(x)\geqslant_w\zeta,[x]_P\notin M_{Pk}\}=\bigcup_{k=1}^{d}\{x|\mu_{\underline{apr}_P(F_k^-)}(x)+\mu_{\underline{apr}_P(F_k^+)}(x)\geqslant\zeta^-+\zeta^+,\mu_{\underline{apr}_P(F_k^+)}(x)-\mu_{\underline{apr}_P(F_k^-)}(x)\geqslant\zeta^+-\zeta^-,[x]_P\notin M_{Pk}\}$. For $\forall x\in T$, put $des_P(x)\to des_{F_k}(x)$ $(k=1,2,\ldots,d)$ into $Rule$.
(11) Output $Rule$.

Remark 13. In Step (8), $des([x])$ represents the antecedent of the rule, and $des_{F_k}(x)$ is the consequent. Such as "If a_3 is 1 Then d is F_2", where, "a_3 is 1" is $des([x])$, "$disF_2$" is $des_{F_k}(x)$.

The time complexity to extract rules is polynomial.

In Step (1), the time complexity for computing a decision partition is $O(|U|^2)$.

In Step (3), the time complexity for computing a significance is $O(|C_j\ \||U|^2)$, then the time complexity of computing $sig_{C_j}^D(c)$ for $\forall c\in C_j$ is $O(|C_j|^2|U|^2)$. The time complexity to choose the minimum of significances is $O(|C_j|)$. In Steps (3) – (5), since$|C|-1$ is the maximum value for the circle times, the time complexity to construct P_j is $\sum_{j=1}^{|C|-1}(O(|C_j|^2|U|^2)+O(|C_j|))=\sum_{j=1}^{|C|-1}O(|C_j|^2|U|^2)+\sum_{j=1}^{|C|-1}O(|C_j|)=O(|C|^2|U|^2+(|C|-1)^2|U|^2+(|C|-2)^2|U|^2\cdots+2^2|U|^2)+O(|C|+(|C|-1)+\cdots+2)=$

$$O\left(\left(\frac{1}{6}(2|C|^3+3|C|^2+|C|)-1\right)|U|^2\right)+O\left(\frac{1}{2}(|C|^2+|C|-2)\right)$$
$$=O(|C|^3|U|^2)$$

In Step (7), the time complexity for computing $\underline{apr}_P(F_k)$, M_{Pk} is $O(|C||U|^2)$.

In Step (8), the time complexity for putting each decision rule into rule base is $O(|M_{Pk}|)$.

In Step (10), the time complexity for computing T is $O(|C||U|^2)$.

In Step (11), the time complexity is $O(|U|)$.

In Steps (7) – (9), d is the circle times. Therefore, the time complexity of the algorithm MRBBA is $O(|U|^2)+ O(|C|^3|U|^2)+\sum_{k=1}^{d}(O(|C||U|^2)+O(|M_{Pk}|))+O(|C||U|^2)+O(|U|)=O(|C|^3|U|^2)$.

The time complexity of this algorithm can be reduced to $O(|C|^3|U|\log_2|U|)$ of a classigication is computed using the ranking technique.

Remark 14. The differences between MRBBA and the method based on attribute reduction include two aspects. First, rule extraction is based on a granulation order, so the adverse effects of attribute reduction are excluded as much as possible with avoiding the attribute reduction process; secondly, MRBBA can extract much simpler decision rules based on keeping the approximation precision invariant.

5.3. An example

In order to discuss the application of MRBBA, an example is given. The dataset given in Table 2 is reused. Where $U=\{x_1,x_2,\ldots,x_{10}\}$ is a set of objects, $C=\{a_1,a_2,a_3\}$ is a condition attribute set, d is a decision attribute, separated into three linguistic terms F_1,F_2,F_3 and F_1,F_2,F_3 are interval-valued fuzzy sets.

According to MRBBA, granulation order is constructed firstly. Compute the significances of a_1, a_2, a_3 with regard to D, respectively. For $C_1=C=\{a_1, a_2, a_3\}$, one can obtain $sig^D_{C_1}(a_1)=0,\ sig^D_{C_1}(a_2)=\frac{335}{2000},\ sig^D_{C_1}(a_3)=\frac{9}{200}$, thus $c_0=a_1$, $C_2=C_1-\{a_1\}=\{a_2, a_3\}$, $P_2=\{C_1, C_2\}$.

For C_2, one can obtain $sig^D_{C_2}(a_2)=\frac{335}{2000},\ sig^D_{C_2}(a_3)=\frac{565}{2000}$. Since $sig^D_{C_2}(a_2)<sig^D_{C_2}(a_3)$, thus $C_3=C_2-\{a_2\}=\{a_3\}$, $P_3=\{C_1, C_2, C_3\}=\{\{a_1, a_2, a_3\}, \{a_2, a_3\}, \{a_3\}\}$ and $P=P_3$.

Let $\zeta=[0.5,0.6]$. From the definition of the backward approximation, one can obtain

$$\mu_{\underline{apr}_P(F_1)}(x_1)=[0.45,0.7],\quad \mu_{\underline{apr}_P(F_1)}(x_2)=[0.05,0.1],\quad \mu_{\underline{apr}_P(F_1)}(x_3)=[0.45,0.7],\quad \mu_{\underline{apr}_P(F_1)}(x_4)=[0.15,0.2],$$
$$\mu_{\underline{apr}_P(F_1)}(x_5)=[0.05,0.1],\quad \mu_{\underline{apr}_P(F_1)}(x_6)=[0.1,0.2],\quad \mu_{\underline{apr}_P(F_1)}(x_7)=[0.05,0.1],\quad \mu_{\underline{apr}_P(F_1)}(x_8)=[0.05,0.1],$$
$$\mu_{\underline{apr}_P(F_1)}(x_9)=[0.45,0.7],\quad \mu_{\underline{apr}_P(F_1)}(x_{10})=[0.05,0.1];$$
$$M_{PF_1}=\emptyset;$$
$$\mu_{\underline{apr}_P(F_2)}(x_1)=[0.15,0.2],\quad \mu_{\underline{apr}_P(F_2)}(x_2)=[0.5\vee0.5\vee0.5,0.7\vee0.7\vee0.7],\quad \mu_{\underline{apr}_P(F_2)}(x_3)=[0.15,0.2],$$
$$\mu_{\underline{apr}_P(F_2)}(x_4)=[0.5\vee0.5,0.8\vee0.8],\quad \mu_{\underline{apr}_P(F_2)}(x_5)=[0.2,0.3],\quad \mu_{\underline{apr}_P(F_2)}(x_6)=[0.35,0.5],\quad \mu_{\underline{apr}_P(F_2)}(x_7)=[0.5\vee0.5\vee0.5,0.7\vee0.7\vee0.7],$$
$$\mu_{\underline{apr}_P(F_2)}(x_8)=[0.2,0.3],\quad \mu_{\underline{apr}_P(F_2)}(x_9)=[0.15,0.2],\quad \mu_{\underline{apr}_P(F_2)}(x_{10})=[0.5\vee0.5\vee0.5,0.7\vee0.7\vee0.7];$$
$$M_{PF_2}=\{\{x_2,x_7,x_{10}\},\{x_4\}\}.$$
$$\mu_{\underline{apr}_P(F_3)}(x_1)=[0.1,0.2],\quad \mu_{\underline{apr}_P(F_3)}(x_2)=[0.05,0.2],\quad \mu_{\underline{apr}_P(F_3)}(x_3)=[0.1,0.2],\quad \mu_{\underline{apr}_P(F_3)}(x_4)=[0.2,0.3],$$
$$\mu_{\underline{apr}_P(F_3)}(x_5)=[0.5\vee0.5\vee0.5,0.6\vee0.6\vee0.6],\quad \mu_{\underline{apr}_P(F_3)}(x_6)=[1.0\vee1.0\vee1.0,1.0\vee1.0\vee1.0],\quad \mu_{\underline{apr}_P(F_3)}(x_7)=[0.05,0.2],$$
$$\mu_{\underline{apr}_P(F_3)}(x_8)=[0.5\vee0.5\vee0.5,0.6\vee0.6\vee0.6],\quad \mu_{\underline{apr}_P(F_3)}(x_9)=[0.1,0.2],\quad \mu_{\underline{apr}_P(F_3)}(x_{10})=[0.05,0.2];$$
$$M_{PF_3}=\{\{x_5,x_8\},\{x_6\}\}.$$

Moreover, for $x\in\{x_1, x_3, x_9\}$, $\mu_{\underline{apr}_P(F_1^-)}(x)+\mu_{\underline{apr}_P(F_1^+)}(x)=0.45+0.7\geqslant\zeta^-+\zeta^+=0.5+0.6$, $\mu_{\underline{apr}_P(F_1^+)}(x)-\mu_{\underline{apr}_P(F_1^-)}(x)=0.7-0.45\geqslant\zeta^+-\zeta^-=0.6-0.5$ and $[x_1]_P\notin M_{PF_k}, [x_3]_P\notin M_{PF_k},\ [x_9]_P\notin M_{PF_k}$, $k=1, 2, 3$. Therefore $T=\{x_1,x_3,x_9\}$. Decision rules can be obtained as follows:

Rule $=\{r_1\prime: des_{\{a_3\}}(x_2,x_7,x_{10})\to des_{F_2}(x)$, namely: if a_3 is 1 Then d is F_2 and $\mu_{\underline{apr}_P(F_2)}(x)\geqslant[0.5,0.7]$; $r_2\prime: des_{\{a_2,a_3\}}(x_4)\to des_{F_2}(x)$, namely: If a_2 is 2 anda_3 is 3 Then d is F_2 and $\mu_{\underline{apr}_P(F_2)}(x)\geqslant[0.5,0.8]$; $r_3': des_{\{a_3\}}(x_5,x_8)\to des_{F_3}(x)$, namely: If a_3 is 4 Then d is F_3 and $\mu_{\underline{apr}_P(F_3)}(x)\geqslant[0.5,0.6]$; $r_4': des_{\{a_3\}}(x_6)\to des_{F_3}(x)$, namely: if a_3 is 2 Then d is F_3 and $\mu_{\underline{apr}_P(F_3)}(x)=[1.0,1.0]$;$r_5\prime: des_{\{a_1,a_2,a_3\}}(x_1,x_3,x_9)\to des_{F_1}(x)$, namely: If a_1 is 2, a_2 is 1 and a_3 is 3. Then d is F_1 and $\mu_{\underline{apr}_P(F_1)}(x)\geqslant[0.45,0.7]\}$.

By comparing r_1', r_2', r_3', r_4' and r_5' with r_1, r_4, r_2, r_3 and r_5 in Section 4.3 respectively, one can easy see that the decision rules extracted from MRBFA and MRBBA are the same.

6. Experiments analysis

6.1. Comparison with the rule extraction method in Gong et al. (2008)

Methods of fuzzy rule extraction based on rough set theory are rarely reported in interval-valued fuzzy information systems. A representative work is found in Gong et al. (2008). The essential of the method in Gong et al. (2008) is to classify each object to corresponding decision classes according to its maximal membership denoted by a fuzzy interval. However, suppose that the condition attribute set includes m attributes, then the antecedent of the rule must include m conditions, overfull conditions may reduce the classification accuracy and the applicability of the rules. Moreover, two memberships denoted by fuzzy intervals are incomparable when one interval is nested in the other, then decision rules cannot be generated in this case.

In order to compare MRBFA and MRBBA with the method in Gong et al. (2008), the dataset given in Table 2 is reused. Using the knowledge discovery method in Gong et al. (2008), the decision rules are generated as follows (Gong et al., 2008):

Rule 1: IF $(a_1,a_2,a_3)=(2, 1, 3)$ THEN d is F_1, the precision of the decision is [0.45,0.7];
Rule 2: IF $(a_1,a_2,a_3)=(3,2,1)$ THEN d is F_2, the precision of the decision is [0.5,0.7];
Rule 3: IF $(a_1,a_2,a_3)=(2,2,3)$ THEN d is F_2, the precision of the decision is [0.5,0.8];
Rule 4: IF $(a_1,a_2,a_3)=(1,1,4)$ THEN d is F_3, the precision of the decision is [0.5,0.6];
Rule 5: IF $(a_1,a_2,a_3)=(1,1,2)$ THEN d is F_3, the precision of the decision is [1.0,1.0].

By comparing the above rules with the ones in Sections 4.3 and 5.3, one can easy see that Rule 1 is the same as Rule r_5 and r'_5, Rule 2 corresponds to r_1 and r'_1, Rule 3 corresponds to r_4 and r'_2, Rule 4 corresponds to r_2 and r'_3, Rule 5 corresponds to r_3 and r'_4. However, conditions satisfied by Rules 2, 3, 4, 5 are far more than the ones of rules generated by MRBFA and MRBBA. Therefore, rules generated by MRBFA and MRBBA possess more broad applicability. Moreover, for MRBFA and MRBBA, decision rules can be generated when an interval is nested in the other due to Step (5) in MRBFA and Step (10) in MRBBA.

6.2. A practical application

In this section, rules are extracted using MRBFA and MRBBA, respectively. Considering that MRBFA, MRBBA mainly deal with the information systems with both crisp condition and interval-valued fuzzy decision attributes, but there is almost no any dataset satisfies the above conditions in existing public database, we need to perform some pretreatments to existing public dataset. We select the dataset "housing" from UCI Machine Learning database (http://www.ics.uci.edu/~mlearn/MLRepository.html) to implement our proposed methods and extract fuzzy rules (experiment is performed on a 400 MHz Pentium Server with 512 MB of memory, running windows xp. Algorithms were coded in Matlab7.1). There are 12 continuous condition attributes, 1 binary-valued condition attribute "CHAS" and a continuous class attribute "MEDV" in this database. There is no missing attribute value. Pretreatments include the discretization of condition attributes, the fuzzification of decision attributes and the conversion of a fuzzy set to an interval-valued fuzzy set.

The discretization of condition attributes is implemented by equal frequency scaler (Wang, 2001). A simple algorithm (Yuan & Shaw, 1995) is used to generate a triangular membership function defined as follows:

$$T_1(x) = \begin{cases} 1, & x \leqslant m_1, \\ (m_2 - x)/(m_2 - m_1), & m_1 < x < m_2, \\ 0, & m_2 \leqslant x, \end{cases}$$

$$T_k(x) = \begin{cases} 1, & x \geqslant m_k, \\ (x - m_{k-1})/(m_k - m_{k-1}), & m_{k-1} < x < m_k, \\ 0, & x \leqslant m_{k-1}, \end{cases}$$

$$T_i(x) = \begin{cases} 0, & x \geqslant m_{i+1}, \\ (m_{i+1} - x)/(m_{i+1} - m_i), & m_i \leqslant x \leqslant m_{i+1}, \\ (x - m_{i-1})/(m_i - m_{i-1}), & m_{i-1} < x < m_i. \\ 0, & x \leqslant m_{i-1} \end{cases}.$$

The slopes of the triangular membership functions are selected such that adjacent membership functions cross at the membership value 0.5. In this case, the only parameter to be determined is the set of $k$centers $M = \{m_i, i = 1,2,\ldots,k\}$. The centers m_i can be calculated using Kohonen's feature-maps algorithm (Kohonen, 1988).

Decision attributes have been fuzzified for the dataset "housing". A construction theorem is used to construct an interval-valued fuzzy set from a fuzzy set (Liu, 2000). Then an interval-valued fuzzy information system can be obtained. There are 506 objects, 13 discrete condition attributes and an interval-valued fuzzy decision attribute that includes three interval-valued attribute values, every attribute value is an interval-valued fuzzy set

Table 3
Classification accuracies of MRBFA, MRBBA and KD.

The number of objects in training set	The number of objects in test set	Method	Training accuracy (%)	Testing accuracy (%)
506	506	MRBFA($\eta^- = 0.5$, $\eta^+ = 0.55$)	0.65834	0.65834
		MRBFA($\eta^- = 0.5$, $\eta^+ = 0.52$)	0.68415	0.68415
		MRBFA($\eta^- = 0.5$, $\eta^+ = 0.5$)	0.73577	0.73577
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.55$)	0.65333	0.65333
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.52$)	0.67756	0.67756
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.5$)	0.72369	0.72369
		KD	0.99290	0.99290
425	81	MRBFA($\eta^- = 0.5$, $\eta^+ = 0.55$)	0.69529	0.52383
		MRBFA($\eta^- = 0.5$, $\eta^+ = 0.52$)	0.72647	0.58617
		MRBFA($\eta^- = 0.5$, $\eta^+ = 0.5$)	0.75824	0.66852
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.55$)	0.68934	0.52870
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.52$)	0.71894	0.59852
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.5$)	0.75647	0.67149
		KD	0.97754	0.34568
350	156	MRBFA($\eta^- = 0.5$, $\eta^+ = 0.55$)	0.69957	0.53923
		MRBFA($\eta^- = 0.5$, $\eta^+ = 0.52$)	0.75571	0.62564
		MRBFA($\eta^- = 0.5$, $\eta^+ = 0.5$)	0.81814	0.68564
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.55$)	0.69637	0.53983
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.52$)	0.75429	0.62282
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.5$)	0.81286	0.69154
		KD	0.96882	0.33974
250	256	MRBFA($\eta^- = 0.5$, $\eta^+ = 0.55$)	0.70600	0.55438
		MRBFA($\eta^- = 0.5$, $\eta^+ = 0.52$)	0.75800	0.62438
		MRBFA($\eta^- = 0.5$, $\eta^+ = 0.5$)	0.82200	0.70609
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.55$)	0.69300	0.55007
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.52$)	0.75400	0.62688
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.5$)	0.81643	0.70134
		KD	0.96800	0.29688
150	356	MRBFA($\eta^- = 0.5$, $\eta^+ = 0.55$)	0.78333	0.59157
		MRBFA($\eta^- = 0.5$, $\eta^+ = 0.52$)	0.81333	0.65562
		MRBFA($\eta^- = 0.5$, $\eta^+ = 0.5$)	0.87333	0.73124
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.55$)	0.78311	0.59361
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.52$)	0.80667	0.65932
		MRBBA($\zeta^- = 0.5$, $\zeta^+ = 0.5$)	0.88903	0.74825
		KD	0.97333	0.22753

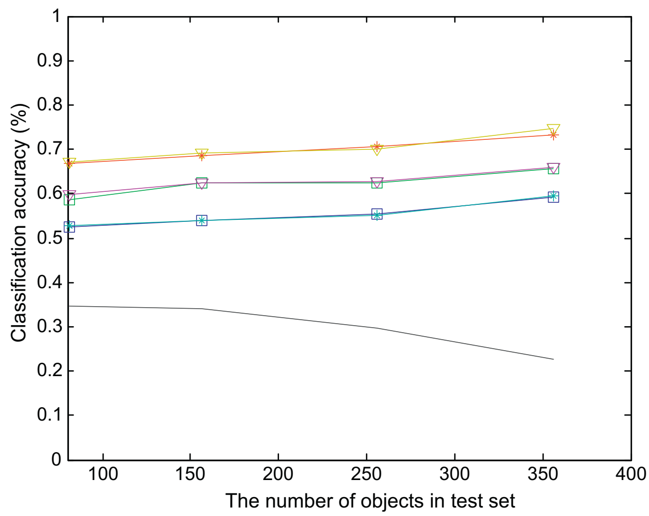

Fig. 1. Testing accuracy. —□—: MRBFA($\eta^- = 0.5$, $\eta^+ = 0.55$); —□—: MRBFA ($\eta^- = 0.5$, $\eta^+ = 0.52$); —∗—: MRBFA($\eta^- = 0.5$, $\eta^+ = 0.5$); —∗—: MRBBA($\xi^- = 0.5$, $\xi^+ = 0.55$); —▽—: MRBBA($\xi^- = 0.5$, $\xi^+ = 0.52$); —▽—: MRBBA($\xi^- = 0.5$, $\xi^+ = 0.5$); —: KD.

on the data set. Rule extraction on interval-valued fuzzy data is then performed using the three methods: MRBFA and MRBBA, the algorithms proposed in Gong et al. (2008) (denoted as KD).

The classification accuracy (including training accuracy and testing accuracy) is used for the evaluation of the three algorithms. The selected dataset is firstly divided into two parts: the training set composed of some randomly chosen samples, and the test set composed of the remainder. Table 3 enumerated part training accuracies and testing accuracies corresponding to different sizes of data sets and different thresholds. Thresholds η and ζ can be considered as the parameters to control the granularity of fuzzy rules. We take the values of η and ζ from [0.5,0.5] to [1.0,1.0] with Step 0.01. The classification accuracies vary with the thresholds. Generally, $[0.5,0.5] \sim [0.5,0.7]$ is a candidate range for η and ζ, where both training and testing accuracies obtain good classification performance. Table 3 shows that two facts: (1) the classification accuracies of MRBFA and MRBBA are almost equivalent; (2)the training accuracies of KD are more than that of MRBFA and MRBBA, however, the testing accuracies of KD are far less than that of MRBFA and MRBBA. Such as, for the data set that includes 150 training samples and 356 test samples, the average classification accuracy obtained for MRBFA ($\eta^- = 0.5$, $\eta^+ = 0.5$) are 87.333% (training accuracy) and 73.124% (testing accuracy), for MRBBA ($\zeta^- = 0.5$, $\zeta^+ = 0.5$) are 88.903% and 74.825%. For KD, the accuracy is 97.333% using the training data, 22.753% for the test data. The objective of learning is to extract rules that can be used to predict the logical implication as accurate as possible when applied to new examples, so testing accuracy is an important criterion to evaluate a rule extraction method. Though the rules obtained by KD have better classification accuracies with reference to the training data set, their generalization ability are rather low since perfect match of condition attribute values is generally difficult. It can be seen from Table 3 that the testing accuracies of KD are almost under 35% (only when the test set is the same as the training set, the testing accuracy is 99.29%, which is equivalent to the training accuracy), it's unpractical in application. Taking one with another, MRBFA and MRBBA outperforms KD. More intuitionistic comparisons can be found in Fig. 1.

7. Conclusions

This paper presents two fuzzy rule extraction methods for interval-valued fuzzy information systems. The main advantages of the methods cover four aspects: firstly, rule extraction is based on a granulation order, so the adverse effects of attribute reduction are excluded as much as possible; secondly, for MRBFA, computational consumption can be reduced effectively as the domain gradually narrows; thirdly, the applicability of the extracted rules by using MRBFA and MRBBA is more broader than the ones obtained by KD; finally, rules can still be generated when one interval is nested in the other. The examples explain the operation mechanism of the rule extraction algorithms and the experiment results show that the two algorithms are reasonable and effective.

Acknowledgments

This work has been supported by the National Natural Science Foundation of China (Nos. 60475019 and 60775036), the research fund for the doctoral program of higher education (No. 20060247039).

References

Asharafa, S., & Murty, M. N. (2003). An adaptive rough fuzzy single pass algorithm for clustering large data sets. *Pattern Recognition, 36*, 3015–3018.

Asharafa, S., & Murty, M. N. (2004). A rough fuzzy approach to web usage categorization. *Fuzzy Sets and Systems, 148*, 119–129.

Bhatt, R. B., & Gopal, M. (2005). On the compact computational domain of fuzzy-rough sets. *Pattern Recognition Letters, 26*, 1632–1640.

Bhatt, R. B., & Gopal, M. (2005). On fuzzy-rough sets approach to feature selection. *Pattern Recognition Letters, 26*, 965–975.

Deng, T. Q., Chen, Y. M., Xu, W. L., & Dai, Q. H. (2007). A novel approach to fuzzy rough sets based on a fuzzy covering. *Information Sciences, 177*, 2308–2326.

Dubois, D., & Prade, H. (1987). Twofold fuzzy sets and rough sets-some issues in knowledge representation. *Fuzzy Sets and Systems, 23*, 3–18.

Dubois, D., & Prade, H. (1990). Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems, 17*(2–3), 191–209.

Gong, Z. T., Sun, B. Z., & Chen, D. G. (2008). Rough set theory for the interval-valued fuzzy information systems. *Information Sciences, 178*, 1968–1985.

Gorzalczany, B. (1988). Interval-valued fuzzy controller based on verbal modal of object. *Fuzzy Sets and Systems, 28*, 45–53.

Greco, S., Inuiguchi, M., & Slowinski, R. (2006). Fuzzy rough sets and multiple-premise gradual decision rules. *International Journal of Approximate Reasoning, 41*, 179–211.

Grzymala-Busse, J. W., & Hu, M. (2001). A comparison of several approaches to missing attribute values in data mining. In W. Ziarko & Y. Yao (Eds.). *RSCTC'00, Lecture Notes in Artificial Intelligence* (Vol. 2005, pp. 378–385). Berlin: Springer.

Grzymala-Busse, J. W., & Grzymala-Busse, W. J. (2007). An experimental comparison of three rough set approaches to missing attribute values. In Peters, J.F. et al. (Eds.), *Transactions on rough sets VI. LNCS* (Vol. 4374, pp. 31–50).

Hu, Q. H., & Yu, D. R. (2004). Entropies of fuzzy indiscernibility relation and its operations. *International Journal of Uncertainty Fuzziness Knowledge Based System, 12*(5), 575–589.

Hu, Q. H., Yu, D. R., & Xie, Z. X. (2006). Information-preserving hybrid data reduction based on fuzzy-rough techniques. *Pattern Recognition Letters, 27*, 414–423.

Hu, H., Xie, Z. X., & Yu, D. R. (2007). Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. *Pattern Recognition, 40*, 3509–3521.

Jensen, R., & Shen, Q. (2004). Fuzzy-rough attribute reduction with application to web categorization. *Fuzzy Sets and Systems, 141*, 469–485.

Jensen, R., & Shen, Q. (2007). Fuzzy-rough sets assisted attribute selection. *IEEE Transactions on Fuzzy Systems, 15*(1), 73–89.

Jiang, J. S., Wu, C. X., & Chen, D. G. (2005). The product structure of fuzzy rough sets on a group and the rough T-fuzzy group. *Information Sciences, 175*, 97–107.

Kohonen, T. (1988). *Self-organization and associative memory*. Berlin: Springer.

Kononenko, I., Bratko, I., & Roskar, E. (1984). *Experiments in automatic learning of medical diagnostic rules*. Technical Report, Jozef Stefan Institute, Ljubljana, Yugoslavia.

Liang, J. Y., Qian, Y. H., Chu, C. Y., Li, D. Y., & Wang, J. H. (2005). Rough set approximation based on dynamic granulation. In Slezak, D. et al. (Eds.), *RSFDGrC 2005. LNAI* (Vol. 3641, pp. 701–708).

Liu, H. W. (2000). Generated construction of interval valued fuzzy sets. *Journal of Shandong University of Technology, 30*, 547–551.

Miao, D. Q., Li, D. G., & Fan, S. D. (2005). Fuzzy rough set and its improvement. In *Proceedings of the first IEEE international conference on granular computing* (pp. 1247–1251). Beijing, China: IEEE Press.

Moshkov, M., Skowron, A., & Suraj, Z. (2008). Maximal consistent extensions of information systems relative to their theories. *Information Sciences, 178*, 2600–2620.

Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Sciences, 11*, 341–356.

Qian, Y. H. et al. (2008). Converse approximation and rule extraction from decision tables in rough set theory. *Computers and Mathematics with Applications, 55*, 1754–1765.

Qin, H. N. (2005). Completing the default in non-completed data tables using valued tolerance relation. *Journal of Wuyi University, 19*, 42–44.

Radzikowska, A. M., & Kerre, E. E. (2002). A comparative study of fuzzy rough sets. *Fuzzy Sets and Systems, 126*, 137–155.

Rajen, B. B., & Gopal, M. (2005). On fuzzy-rough sets approach to feature selection. *Pattern Recognition Letters, 26*, 965–975.

Richard, J., & Qiang, Sh. (2002). Fuzzy-rough sets for descriptive dimensionality reduction. In *Proceedings of the 2002 IEEE international conference fuzzy systems, FUZZ-IEEE'02* (pp. 29–34).

Sankar, K. P. (2004). Soft data mining, computational theory of perceptions, and rough-fuzzy approach. *Information Sciences, 163*, 5–12.

Shen, Q., & Chouchoulas, A. (2002). A rough-fuzzy approach for generating classification rules. *Pattern Recognition, 35*, 2425–2438.

Shen, Q., & Jensen, R. (2004). Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring. *Pattern Recognition, 37*, 1351–1363.

Tsang, G. C. Y., Chen, D. G., Tsang, E. C. C., Lee, J. W. T., & Yeung, D. S. (2005). On attribute reduction with fuzzy rough sets. In *Proceedings of SMC'05* (pp. 2775–2780).

Tsang, E. C. C., Chen, D. G., Yeung, D. S., Wang, X. Z., & Lee, J. W. T. (2008). Attribute reduction using fuzzy rough sets. *IEEE Transaction on Fuzzy Systems., 16*(5), 1130–1141.

Turksen, B. (1986). Interval-valued fuzzy sets based on normal forms. *Fuzzy Sets and Systems, 20*, 191–210.

Wang, G. Y. (2001). *Rough set theory and data mining*. Xi'an: Xi'an Jiaotong University Press.

Wang, X. Z., Ha, Y., & Chen, D. G. (2005). On the reduction of fuzzy rough sets. In *Proceedings of ICMLC'05* (pp. 3174–3178). IEEE Press.

Wang, X. Z., Tsang, E. C. C., Zhao, S. Y., Chen, D. G., & Yeung, D. S. (2007). Learning fuzzy rules from fuzzy samples based on rough set technique. *Information Sciences, 177*, 4493–4514.

Wang, X. Z., Zhai, J. H., & Lu, S. X. (2008). Induction of multiple fuzzy decision trees based on rough set technique. *Information Sciences, 178*, 3188–3202.

Wang, Y. F. (2003). Mining stock price using fuzzy rough set system. *Expert Systems with Applications, 24*, 13–23.

Yeung, D. S., Chen, D. G., Tsang, E. C. C., Lee, J. W. T., & Wang, X. Z. (2005). On the generalization of fuzzy rough sets. *IEEE Transactions on Fuzzy Systems, 13*, 343–361.

Yuan, Y. F., & Shaw, M. J. (1995). Introduction of fuzzy decision tree. *Fuzzy Sets and Systems, 69*, 125–139.

Zhang, W. X., Wu, W. Z., Liang, J. Y., & Li, D. Y. (2001). *The theory and methodology of rough sets*. Beijing: The Science Press (in Chinese).

Zhao, S. Y., Tsang, E. C. C., Wang, X. Z., Chen, D. G., & Yeung, D. S. (2006). Fuzzy matrix computation for fuzzy information system to reduce attributes. In *Proceedings of ICMLC'06* (pp. 2300–2304). IEEE Press.

Zhou, L., & Wu, W. Z. (2008). On generalized intuitionistic fuzzy rough approximation operators. *Information Sciences, 178*, 2448–2465.

Zhu, W., Zhang, W., & Fu, Y. (2004). An incomplete data analysis approach using rough set theory. In *Proceedings of international conference on intelligent mechatronics and automation* (pp. 332–338).